

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 09:46:42 ; Search time 15 Seconds  
(without alignments)  
351.114 Million cell updates/sec

Title: US-09-995-493-52

Perfect score: 179

Sequence: 1 OHNGVLGPYIGKSLTLKP.....ADYTTKATANLYGLNLYRF 179

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued\_Patents\_AA.\*

1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13	7.3	77	2	US-08-472-172-12
2	13	7.3	432	2	US-08-472-172-6
3	13	7.3	455	2	US-08-472-172-4
4	13	7.3	459	2	US-08-472-172-2
5	10	5.6	34	1	US-08-475-989-8
6	10	5.6	34	2	US-08-475-985-8
7	10	5.6	34	2	US-08-472-172-20
8	10	5.6	34	3	US-08-256-839-8
9	10	5.6	35	2	US-08-475-989-14
10	10	5.6	35	2	US-08-475-985-14
11	10	5.6	35	2	US-08-472-172-26
12	10	5.6	35	3	US-08-256-839-14
13	10	5.6	38	1	US-08-475-989-13
14	10	5.6	38	2	US-08-475-985-13
15	10	5.6	38	2	US-08-472-172-25
16	10	5.6	38	3	US-08-256-839-13
17	10	5.6	53	1	US-08-475-989-48
18	10	5.6	53	2	US-08-475-985-48
19	10	5.6	53	3	US-08-256-839-48
20	10	5.6	54	1	US-08-475-989-49
21	10	5.6	54	2	US-08-475-985-49
22	10	5.6	54	3	US-08-256-839-49
23	10	5.6	55	1	US-08-475-989-47
24	10	5.6	55	2	US-08-475-985-47
25	10	5.6	55	3	US-08-256-839-47
26	7	3.9	12	4	US-09-347-926-9
27	7	3.9	59	4	US-09-300-672-7

28	7	3.9	61	2	US-08-949-603-11	Sequence 11, Appl
29	7	3.9	61	2	US-08-706-270A-11	Sequence 11, Appl
30	7	3.9	61	2	US-08-949-580-11	Sequence 11, Appl
31	7	3.9	61	2	US-08-950-172A-11	Sequence 11, Appl
32	7	3.9	61	4	US-09-198-119C-11	Sequence 11, Appl
33	7	3.9	68	3	US-08-912-272-28	Sequence 28, Appl
34	7	3.9	68	4	US-09-026-039-28	Sequence 28, Appl
35	7	3.9	69	3	US-08-912-272-18	Sequence 18, Appl
36	7	3.9	69	4	US-09-026-039-18	Sequence 18, Appl
37	7	3.9	466	2	US-08-726-136-28	Sequence 28, Appl
38	7	3.9	466	4	US-09-103-434-28	Sequence 28, Appl
39	7	3.9	466	4	US-09-687-594-28	Sequence 28, Appl
40	6	3.4	15	1	US-08-056-200-60	Sequence 60, Appl
41	6	3.4	15	1	US-08-056-200-61	Sequence 61, Appl
42	6	3.4	15	2	US-08-800-644-60	Sequence 60, Appl
43	6	3.4	15	2	US-08-800-644-61	Sequence 61, Appl
44	6	3.4	18	3	US-08-912-272-7	Sequence 7, Appl
45	6	3.4	18	4	US-09-026-039-7	Sequence 7, Appl
46	6	3.4	20	3	US-08-467-023-40	Sequence 40, Appl
47	6	3.4	20	3	US-08-467-023-41	Sequence 41, Appl
48	6	3.4	21	4	US-09-288-143-181	Sequence 181, Appl
49	6	3.4	22	4	US-08-973-801-15	Sequence 15, Appl
50	6	3.4	26	1	US-08-484-635-91	Sequence 91, Appl
51	6	3.4	26	2	US-08-484-631-91	Sequence 91, Appl
52	6	3.4	26	2	US-08-827-570-91	Sequence 91, Appl
53	6	3.4	30	3	US-08-256-747C-44	Sequence 44, Appl
54	6	3.4	30	3	US-08-467-023-68	Sequence 68, Appl
55	6	3.4	30	3	US-08-467-023-75	Sequence 75, Appl
56	6	3.4	30	4	US-08-467-023-75	Sequence 75, Appl
57	6	3.4	32	1	US-08-834-130A-44	Sequence 44, Appl
58	6	3.4	32	1	US-08-473-989-10	Sequence 10, Appl
59	6	3.4	32	2	US-08-473-985-10	Sequence 10, Appl
60	6	3.4	32	2	US-08-472-172-22	Sequence 22, Appl
61	6	3.4	32	3	US-08-256-839-10	Sequence 10, Appl
62	6	3.4	30	3	US-08-467-023-63	Sequence 63, Appl
63	6	3.4	53	3	US-08-912-272-12	Sequence 12, Appl
64	6	3.4	53	4	US-09-026-039-12	Sequence 12, Appl
65	6	3.4	60	4	US-09-300-672-9	Sequence 9, Appl
66	6	3.4	68	2	US-08-700-152A-2	Sequence 2, Appl
67	6	3.4	68	3	US-08-912-272-5	Sequence 5, Appl
68	6	3.4	68	4	US-09-026-039-5	Sequence 5, Appl
69	6	3.4	69	3	US-08-912-272-17	Sequence 17, Appl
70	6	3.4	69	4	US-09-026-039-17	Sequence 17, Appl
71	6	3.4	149	4	US-09-227-357-134	Sequence 134, Appl
72	6	3.4	175	4	US-09-288-143-173	Sequence 173, Appl
73	6	3.4	245	4	US-09-413-814-6	Sequence 6, Appl
74	6	3.4	253	4	US-09-288-143-84	Sequence 84, Appl
75	6	3.4	262	3	US-08-946-914-14	Sequence 14, Appl
76	6	3.4	262	4	US-09-656-450-14	Sequence 14, Appl
77	6	3.4	264	1	US-08-562-311-4	Sequence 4, Appl
78	6	3.4	276	4	US-08-961-083-134	Sequence 134, Appl
79	6	3.4	282	4	US-08-288-143-172	Sequence 172, Appl
80	6	3.4	297	4	US-08-765-907A-3	Sequence 3, Appl
81	6	3.4	299	1	US-08-169-950-3	Sequence 3, Appl
82	6	3.4	333	4	US-09-338-420-4	Sequence 4, Appl
83	6	3.4	359	4	US-09-134-001C-4842	Sequence 4842, Ap
84	6	3.4	374	3	US-08-467-023-2	Sequence 2, Appl
85	6	3.4	386	4	US-09-338-420-2	Sequence 2, Appl
86	6	3.4	393	2	US-08-933-750C-11	Sequence 11, Appl
87	6	3.4	393	4	US-09-234-613-11	Sequence 11, Appl
88	6	3.4	411	4	US-09-188-930-302	Sequence 302, Appl
89	6	3.4	432	1	US-08-700-152A-4	Sequence 4, Appl
90	6	3.4	446	2	US-07-627-539G-7	Sequence 7, Appl
91	6	3.4	452	3	US-09-059-522-3	Sequence 3, Appl
92	6	3.4	452	4	US-09-382-027-3	Sequence 3, Appl
93	6	3.4	468	1	US-07-627-539G-2	Sequence 2, Appl
94	6	3.4	490	4	US-09-252-149B-26	Sequence 26, Appl
95	6	3.4	496	4	US-08-833-876-2	Sequence 2, Appl
96	6	3.4	496	4	US-09-483-054-2	Sequence 2, Appl
97	6	3.4	503	4	US-09-134-001C-4214	Sequence 4214, Ap
98	6	3.4	507	4	US-09-134-001C-5071	Sequence 5071, Ap
99	6	3.4	512	2	US-08-828-596-2	Sequence 2, Appl
100	6	3.4	512	4	US-09-360-197-6	Sequence 6, Appl

101	6	3.4	524	4	US-09-134-001C-4315	Sequence 4315, Ap	174	6	3.4	908	2	US-08-478-435-94	Sequence 94, Appl
102	6	3.4	533	4	US-09-350-268-3	Sequence 3, Appli	175	6	3.4	908	2	US-08-337-483-94	Sequence 94, Appl
103	6	3.4	533	4	US-09-558-679-3	Sequence 3, Appli	176	6	3.4	908	2	US-08-478-373-94	Sequence 94, Appl
104	6	3.4	544	1	US-08-387-156-10	Sequence 10, Appl	177	6	3.4	908	3	US-08-474-671-94	Sequence 94, Appl
105	6	3.4	544	2	US-08-694-865-10	Sequence 10, Appl	178	6	3.4	908	3	US-08-483-577A-94	Sequence 94, Appl
106	6	3.4	544	2	US-08-878-748-10	Sequence 10, Appl	179	6	3.4	908	3	US-08-448-194-4	Sequence 4, Appli
107	6	3.4	544	3	US-09-124-491-10	Sequence 10, Appl	180	6	3.4	908	3	US-08-613-009A-16	Sequence 16, Appl
108	6	3.4	546	1	US-08-487-890A-99	Sequence 99, Appl	181	6	3.4	908	4	US-08-897-438-94	Sequence 94, Appl
109	6	3.4	546	2	US-08-478-435-99	Sequence 99, Appl	182	6	3.4	908	4	US-08-867-921-4	Sequence 4, Appli
110	6	3.4	546	2	US-08-337-483-99	Sequence 99, Appl	183	6	3.4	908	4	US-08-637-654-94	Sequence 94, Appl
111	6	3.4	546	2	US-08-478-373-99	Sequence 99, Appl	184	6	3.4	908	4	US-08-549-518-94	Sequence 94, Appl
112	6	3.4	546	3	US-08-474-671-99	Sequence 99, Appl	185	6	3.4	908	4	US-08-778-570B-22	Sequence 22, Appl
113	6	3.4	546	3	US-08-483-577A-99	Sequence 99, Appl	186	6	3.4	908	4	US-09-059-584-22	Sequence 22, Appl
114	6	3.4	546	4	US-08-897-438-99	Sequence 99, Appl	187	6	3.4	909	2	US-08-363-124A-4	Sequence 4, Appli
115	6	3.4	546	4	US-08-637-654-99	Sequence 99, Appl	188	6	3.4	919	4	US-08-985-916-16	Sequence 16, Appl
116	6	3.4	546	4	US-08-649-518-99	Sequence 99, Appl	189	6	3.4	924	3	US-08-619-812-8	Sequence 8, Appli
117	6	3.4	547	1	US-07-961-522-2	Sequence 2, Appli	190	6	3.4	926	1	US-07-908-253-2	Sequence 2, Appli
118	6	3.4	547	1	US-07-961-522-5	Sequence 5, Appli	191	6	3.4	926	1	US-08-455-970A-2	Sequence 2, Appli
119	6	3.4	547	1	US-08-217-438-2	Sequence 2, Appli	192	6	3.4	926	1	US-08-387-156-6	Sequence 6, Appli
120	6	3.4	547	1	US-08-217-438-6	Sequence 6, Appli	193	6	3.4	926	2	US-08-694-865-6	Sequence 6, Appli
121	6	3.4	547	1	US-08-321-978-2	Sequence 2, Appli	194	6	3.4	926	2	US-08-878-748-6	Sequence 6, Appli
122	6	3.4	547	1	US-08-321-978-5	Sequence 5, Appli	195	6	3.4	926	2	US-08-535-837-2	Sequence 2, Appli
123	6	3.4	547	2	US-08-710-584-2	Sequence 2, Appli	196	6	3.4	926	3	US-09-124-491-6	Sequence 6, Appli
124	6	3.4	547	2	US-08-710-584-5	Sequence 5, Appli	197	6	3.4	926	6	5476657-3	Patent No. 5476657
125	6	3.4	553	4	US-09-422-936-79	Sequence 79, Appl	198	6	3.4	927	3	US-08-895-601-6	Sequence 6, Appli
126	6	3.4	575	1	US-08-403-866-7	Sequence 7, Appli	199	6	3.4	936	1	US-08-455-970A-12	Sequence 12, Appl
127	6	3.4	575	1	US-08-620-605D-2	Sequence 2, Appli	200	6	3.4	943	1	US-08-455-970A-10	Sequence 10, Appl
128	6	3.4	588	2	US-09-005-232A-2	Sequence 2, Appli	201	6	3.4	951	1	US-08-455-970A-14	Sequence 14, Appl
129	6	3.4	600	3	US-08-537-361B-9	Sequence 9, Appli	202	6	3.4	960	4	US-09-422-936-45	Sequence 45, Appl
130	6	3.4	602	3	US-08-990-470A-3	Sequence 3, Appli	203	6	3.4	961	4	US-09-422-936-49	Sequence 49, Appl
131	6	3.4	602	4	US-08-817-707-9	Sequence 9, Appli	204	6	3.4	964	4	US-09-422-936-53	Sequence 53, Appl
132	6	3.4	614	1	US-08-652-207A-2	Sequence 2, Appli	205	6	3.4	977	1	US-08-387-156-8	Sequence 8, Appli
133	6	3.4	621	3	US-09-059-522-1	Sequence 1, Appli	206	6	3.4	977	2	US-08-694-865-8	Sequence 8, Appli
134	6	3.4	621	4	US-09-382-027-1	Sequence 1, Appli	207	6	3.4	977	2	US-08-878-748-8	Sequence 8, Appli
135	6	3.4	622	2	US-08-664-646A-2	Sequence 2, Appli	208	6	3.4	977	3	US-09-124-491-8	Sequence 8, Appli
136	6	3.4	622	2	US-09-066-285-2	Sequence 2, Appli	209	6	3.4	992	1	US-08-127-499A-1	Sequence 1, Appli
137	6	3.4	622	3	US-09-261-006-2	Sequence 2, Appli	210	6	3.4	992	1	US-08-482-847-1	Sequence 1, Appli
138	6	3.4	622	4	US-08-351-088-2	Sequence 2, Appli	211	6	3.4	1018	1	US-08-408-093-6	Sequence 6, Appli
139	6	3.4	622	4	US-09-609-566-2	Sequence 2, Appli	212	6	3.4	1018	1	US-08-408-420A-6	Sequence 6, Appli
140	6	3.4	622	4	US-09-609-570-2	Sequence 2, Appli	213	6	3.4	1018	1	US-08-714-901-6	Sequence 6, Appli
141	6	3.4	628	1	US-08-257-073-9	Sequence 9, Appli	214	6	3.4	1018	3	US-08-040-741-6	Sequence 6, Appli
142	6	3.4	642	2	US-08-245-511-48	Sequence 48, Appl	215	6	3.4	1069	1	US-07-777-715-9	Sequence 9, Appli
143	6	3.4	642	2	US-08-600-993A-48	Sequence 48, Appl	216	6	3.4	1069	1	US-08-170-126-4	Sequence 4, Appli
144	6	3.4	661	1	US-08-514-014-4	Sequence 4, Appli	217	6	3.4	1069	3	US-08-954-418-4	Sequence 4, Appli
145	6	3.4	661	2	US-08-833-823-4	Sequence 4, Appli	218	6	3.4	1076	4	US-09-134-001C-4037	Sequence 4037, Ap
146	6	3.4	662	1	US-07-841-651-4	Sequence 4, Appli	219	6	3.4	1090	4	US-09-346-237-5	Sequence 5, Appli
147	6	3.4	685	5	PCT-US91-09784-4	Sequence 4, Appli	220	6	3.4	1096	4	US-09-346-237-6	Sequence 6, Appli
148	6	3.4	693	1	US-08-056-200-111	Sequence 111, App	221	6	3.4	1098	1	US-07-777-715-7	Sequence 7, Appli
149	6	3.4	693	1	US-08-056-200-112	Sequence 112, App	222	6	3.4	1098	1	US-08-170-126-2	Sequence 2, Appli
150	6	3.4	693	2	US-08-800-644-111	Sequence 111, App	223	6	3.4	1098	3	US-08-954-418-2	Sequence 2, Appli
151	6	3.4	693	2	US-08-800-644-112	Sequence 112, App	224	6	3.4	1334	6	5476657-1	Patent No. 5476657
152	6	3.4	699	2	US-08-694-865-16	Sequence 16, Appl	225	6	3.4	1375	4	US-09-210-361-4	Sequence 4, Appli
153	6	3.4	699	3	US-09-124-491-16	Sequence 16, Appl	226	6	3.4	1375	4	US-09-740-274-4	Sequence 4, Appli
154	6	3.4	771	1	US-09-090-793-8	Sequence 8, Appli	227	6	3.4	1403	1	US-07-908-253-3	Sequence 3, Appli
155	6	3.4	780	1	US-08-375-709-13	Sequence 13, Appl	228	6	3.4	1403	2	US-08-694-865-17	Sequence 17, Appl
156	6	3.4	780	1	US-08-752-929-13	Sequence 13, Appl	229	6	3.4	1403	2	US-08-535-837-3	Sequence 3, Appli
157	6	3.4	790	3	US-08-537-361E-8	Sequence 8, Appli	230	6	3.4	1403	3	US-09-124-491-17	Sequence 17, Appl
158	6	3.4	790	4	US-08-817-707-6	Sequence 6, Appli	231	6	3.4	1438	4	US-09-209-916-1	Sequence 1, Appli
159	6	3.4	790	4	US-08-817-707-8	Sequence 8, Appli	232	6	3.4	1443	2	US-08-670-707A-39	Sequence 39, Appl
160	6	3.4	791	3	US-08-375-361E-4	Sequence 4, Appli	233	6	3.4	1443	4	US-09-037-601-39	Sequence 39, Appl
161	6	3.4	791	4	US-08-817-707-4	Sequence 4, Appli	234	6	3.4	1443	4	US-09-315-179-39	Sequence 39, Appl
162	6	3.4	792	1	US-08-326-670A-2	Sequence 2, Appli	235	6	3.4	1457	4	US-09-523-656-38	Sequence 38, Appl
163	6	3.4	792	3	US-08-537-361E-2	Sequence 2, Appli	236	6	3.4	1471	1	US-08-683-839B-3	Sequence 3, Appli
164	6	3.4	792	3	US-08-537-361E-6	Sequence 6, Appli	237	6	3.4	1475	3	US-09-007-999-2	Sequence 2, Appli
165	6	3.4	792	3	US-08-990-470A-2	Sequence 2, Appli	238	6	3.4	1475	4	US-09-210-361-2	Sequence 2, Appli
166	6	3.4	792	4	US-08-817-707-2	Sequence 2, Appli	239	6	3.4	1475	4	US-09-740-274-2	Sequence 2, Appli
167	6	3.4	844	4	US-09-422-936-47	Sequence 47, Appl	240	6	3.4	1661	2	US-08-982-083-2	Sequence 2, Appli
168	6	3.4	844	4	US-09-422-936-51	Sequence 51, Appl	241	6	3.4	1661	2	US-08-558-107-2	Sequence 2, Appli
169	6	3.4	884	1	US-08-066-167-2	Sequence 2, Appli	242	6	3.4	1661	4	US-09-243-539-2	Sequence 2, Appli
170	6	3.4	886	4	US-09-422-936-77	Sequence 77, Appl	243	6	3.4	2115	4	US-09-324-867-5	Sequence 5, Appli
171	6	3.4	892	4	US-09-422-936-75	Sequence 75, Appl	244	6	3.4	2133	2	US-08-670-707A-37	Sequence 37, Appl
172	6	3.4	899	4	US-09-422-936-71	Sequence 71, Appl	245	6	3.4	2133	4	US-09-037-601-37	Sequence 37, Appl
173	6	3.4	908	1	US-08-487-890A-94	Sequence 94, Appl	246	6	3.4	2133	4	US-09-315-179-37	Sequence 37, Appl

247	6	3.4	2133	4	US-09-523-656-30	Sequence 30, Appl	Sequence 30, Appl	320	5	2.8	15	6	5496924-4	Patent No. 5496924
248	6	3.4	2304	4	US-09-324-867-4	Sequence 4, Appl1	Sequence 4, Appl1	321	5	2.8	15	6	5496924-13	Patent No. 5496924
249	6	3.4	2319	1	US-08-212-133A-8	Sequence 8, Appl1	Sequence 8, Appl1	322	5	2.8	16	3	US-09-100-414B-2	Sequence 2, Appl1
250	6	3.4	2319	1	US-08-474-503-6	Sequence 6, Appl1	Sequence 6, Appl1	323	5	2.8	16	3	US-09-100-409A-61	Sequence 61, Appl1
251	6	3.4	2319	2	US-08-670-707A-6	Sequence 6, Appl1	Sequence 6, Appl1	324	5	2.8	16	4	US-09-303-323-2	Sequence 2, Appl1
252	6	3.4	2319	4	US-09-037-601-6	Sequence 6, Appl1	Sequence 6, Appl1	325	5	2.8	17	4	US-09-265-653-18	Sequence 18, Appl1
253	6	3.4	2319	4	US-09-315-179-6	Sequence 6, Appl1	Sequence 6, Appl1	326	5	2.8	18	2	US-08-746-283-19	Sequence 19, Appl1
254	6	3.4	2319	4	US-09-523-656-28	Sequence 28, Appl1	Sequence 28, Appl1	327	5	2.8	18	2	US-08-746-257A-17	Sequence 17, Appl1
255	6	3.4	2319	5	PCT-US94-13200-6	Sequence 6, Appl1	Sequence 6, Appl1	328	5	2.8	20	1	US-08-190-788A-148	Sequence 148, App
256	6	3.4	2332	1	US-07-864-004B-4	Sequence 4, Appl1	Sequence 4, Appl1	329	5	2.8	20	1	US-08-383-474B-153	Sequence 153, App
257	6	3.4	2332	1	US-08-251-937A-4	Sequence 4, Appl1	Sequence 4, Appl1	330	5	2.8	20	1	US-08-465-391A-148	Sequence 148, App
258	6	3.4	2332	1	US-08-212-133A-2	Sequence 2, Appl1	Sequence 2, Appl1	331	5	2.8	20	2	US-08-482-142-16	Sequence 16, Appl
259	6	3.4	2332	1	US-08-276-594A-2	Sequence 2, Appl1	Sequence 2, Appl1	332	5	2.8	20	2	US-08-482-142-205	Sequence 205, App
260	6	3.4	2332	1	US-08-474-503-2	Sequence 2, Appl1	Sequence 2, Appl1	333	5	2.8	20	2	US-08-464-538B-148	Sequence 148, App
261	6	3.4	2332	2	US-08-670-707A-2	Sequence 2, Appl1	Sequence 2, Appl1	334	5	2.8	20	2	US-08-463-076B-199	Sequence 199, App
262	6	3.4	2332	4	US-09-037-601-2	Sequence 2, Appl1	Sequence 2, Appl1	335	5	2.8	20	2	US-08-348-353-26	Sequence 26, Appl
263	6	3.4	2332	4	US-09-324-867-3	Sequence 3, Appl1	Sequence 3, Appl1	336	5	2.8	20	2	US-08-465-965-26	Sequence 26, Appl
264	6	3.4	2332	4	US-09-315-179-2	Sequence 2, Appl1	Sequence 2, Appl1	337	5	2.8	20	2	US-08-478-572-16	Sequence 16, Appl
265	6	3.4	2332	4	US-09-523-656-2	Sequence 2, Appl1	Sequence 2, Appl1	338	5	2.8	20	2	US-08-478-572-205	Sequence 205, App
266	6	3.4	2332	5	PCT-US93-03275-4	Sequence 4, Appl1	Sequence 4, Appl1	339	5	2.8	20	2	US-08-466-860-29	Sequence 29, Appl
267	6	3.4	2332	5	PCT-US94-13200-2	Sequence 2, Appl1	Sequence 2, Appl1	340	5	2.8	20	3	US-08-465-966-26	Sequence 26, Appl
268	6	3.4	2343	4	US-09-324-867-2	Sequence 2, Appl1	Sequence 2, Appl1	341	5	2.8	20	3	US-08-472-040A-29	Sequence 29, Appl
269	6	3.4	2351	1	US-08-121-202-2	Sequence 2, Appl1	Sequence 2, Appl1	342	5	2.8	20	3	US-08-840-006-4	Sequence 4, Appl1
270	6	3.4	2351	1	US-08-366-851A-2	Sequence 2, Appl1	Sequence 2, Appl1	343	5	2.8	20	4	US-08-276-776-29	Sequence 29, Appl
271	6	3.4	2351	6	5171844-2	Patent No. 5171844	Patent No. 5171844	344	5	2.8	20	4	US-08-471-209-29	Sequence 29, Appl
272	6	3.4	2351	6	5422260-1	Patent No. 5422260	Patent No. 5422260	345	5	2.8	20	4	US-08-484-296-16	Sequence 16, Appl
273	6	3.4	2584	3	US-08-936-135-4	Sequence 4, Appl1	Sequence 4, Appl1	346	5	2.8	20	4	US-08-484-296-205	Sequence 205, App
274	6	3.4	2588	3	US-08-936-135-2	Sequence 2, Appl1	Sequence 2, Appl1	347	5	2.8	20	4	US-08-055-006-9	Sequence 9, Appl1
275	6	3.4	2710	2	US-08-568-459A-12	Sequence 12, Appl	Sequence 12, Appl	348	5	2.8	20	5	PCT-US95-04481-7	Sequence 7, Appl1
276	6	3.4	2710	2	US-08-487-828B-12	Sequence 12, Appl	Sequence 12, Appl	349	5	2.8	21	2	US-08-466-860-30	Sequence 30, Appl
277	6	3.4	2710	2	US-08-210-288-12	Sequence 12, Appl	Sequence 12, Appl	350	5	2.8	21	3	US-08-472-040A-30	Sequence 30, Appl
278	6	3.4	3060	2	US-08-487-828B-14	Sequence 14, Appl	Sequence 14, Appl	351	5	2.8	21	3	US-08-851-843A-175	Sequence 175, App
279	6	3.4	3169	2	US-08-477-451-6	Sequence 6, Appl1	Sequence 6, Appl1	352	5	2.8	21	4	US-08-974-549A-294	Sequence 294, App
280	6	3.4	3218	1	US-08-764-100-27	Sequence 27, Appl	Sequence 27, Appl	353	5	2.8	21	4	US-08-276-776-30	Sequence 30, Appl
281	6	3.4	4551	3	US-09-320-878-1	Sequence 1, Appl1	Sequence 1, Appl1	354	5	2.8	21	4	US-08-471-209-30	Sequence 30, Appl
282	6	3.4	4613	4	US-09-105-537-31	Sequence 31, Appl	Sequence 31, Appl	355	5	2.8	21	4	US-08-854-050-175	Sequence 175, App
283	6	3.4	11877	4	US-09-105-537-6	Sequence 6, Appl1	Sequence 6, Appl1	356	5	2.8	21	4	US-09-430-323-175	Sequence 175, App
284	5	2.8	1370	2	US-08-350-260A-377	Sequence 377, App	Sequence 377, App	357	5	2.8	22	1	US-08-471-780C-4	Sequence 4, Appl1
285	5	2.8	7	4	US-09-139-802-130	Sequence 130, App	Sequence 130, App	358	5	2.8	22	1	US-08-484-635-141	Sequence 141, App
286	5	2.8	8	4	US-08-660-092-125	Sequence 125, App	Sequence 125, App	359	5	2.8	22	1	US-08-467-282B-4	Sequence 4, Appl1
287	5	2.8	9	4	US-09-160-513-125	Sequence 125, App	Sequence 125, App	360	5	2.8	22	2	US-08-484-631-141	Sequence 141, App
288	5	2.8	9	4	US-09-644-600-27	Sequence 27, Appl	Sequence 27, Appl	361	5	2.8	22	2	US-08-471-282A-4	Sequence 4, Appl1
289	5	2.8	9	4	US-09-644-600-35	Sequence 35, Appl	Sequence 35, Appl	362	5	2.8	22	2	US-08-466-710C-4	Sequence 4, Appl1
290	5	2.8	9	4	US-09-644-600-36	Sequence 36, Appl	Sequence 36, Appl	363	5	2.8	22	2	US-08-827-570-141	Sequence 141, App
291	5	2.8	9	4	US-09-644-600-38	Sequence 38, Appl	Sequence 38, Appl	364	5	2.8	22	3	PCT-US92-08094-5	Sequence 5, Appl1
292	5	2.8	10	1	US-08-299-285-21	Sequence 21, Appl	Sequence 21, Appl	365	5	2.8	22	3	US-08-141-324-25	Sequence 25, Appl
293	5	2.8	10	2	US-08-985-126-21	Sequence 21, Appl	Sequence 21, Appl	366	5	2.8	23	1	US-08-541-902-25	Sequence 25, Appl
294	5	2.8	10	3	US-08-840-006-3	Sequence 3, Appl1	Sequence 3, Appl1	367	5	2.8	23	1	US-08-139-862-11	Sequence 11, Appl
295	5	2.8	10	4	US-09-358-020-21	Sequence 21, Appl	Sequence 21, Appl	368	5	2.8	23	1	US-08-139-862-12	Sequence 12, Appl
296	5	2.8	11	6	PCT-US95-11126-21	Patent No. 5196523	Patent No. 5196523	369	5	2.8	24	1	US-08-321-356-7	Sequence 7, Appl1
297	5	2.8	11	6	5196523-14	Patent No. 5196523	Patent No. 5196523	370	5	2.8	24	1	US-08-484-635-70	Sequence 70, Appl
298	5	2.8	11	6	5464750-1	Patent No. 5464750	Patent No. 5464750	371	5	2.8	24	2	US-08-482-142-130	Sequence 130, App
299	5	2.8	13	3	US-08-486-099-174	Sequence 174, App	Sequence 174, App	372	5	2.8	24	2	US-08-484-631-70	Sequence 70, App
300	5	2.8	13	3	US-08-484-223B-174	Sequence 174, App	Sequence 174, App	373	5	2.8	24	2	US-08-478-572-130	Sequence 130, App
301	5	2.8	13	3	US-08-919-597-174	Sequence 174, App	Sequence 174, App	374	5	2.8	24	2	US-08-827-570-70	Sequence 70, Appl
302	5	2.8	13	3	US-08-475-668A-174	Sequence 174, App	Sequence 174, App	375	5	2.8	24	3	US-08-545-196B-59	Sequence 59, Appl
303	5	2.8	13	3	US-08-485-551A-174	Sequence 174, App	Sequence 174, App	376	5	2.8	24	4	US-08-484-296-130	Sequence 130, App
304	5	2.8	13	3	US-08-471-913A-174	Sequence 174, App	Sequence 174, App	377	5	2.8	25	2	US-08-913-475A-6	Sequence 6, Appl1
305	5	2.8	13	3	US-09-101-146-47	Sequence 47, Appl	Sequence 47, Appl	378	5	2.8	25	4	US-08-078-683A-15	Sequence 15, Appl
306	5	2.8	13	4	US-08-485-264A-174	Sequence 174, App	Sequence 174, App	379	5	2.8	28	1	US-08-446-692-38	Sequence 38, Appl
307	5	2.8	13	4	US-09-082-279B-552	Sequence 552, App	Sequence 552, App	380	5	2.8	28	2	US-08-482-142-15	Sequence 15, Appl
308	5	2.8	13	4	US-08-474-349A-174	Sequence 174, App	Sequence 174, App	381	5	2.8	28	2	US-08-482-142-33	Sequence 33, Appl
309	5	2.8	13	4	US-09-315-304B-532	Sequence 532, App	Sequence 532, App	382	5	2.8	28	2	US-08-482-142-126	Sequence 126, App
310	5	2.8	13	6	5496924-54	Patent No. 5496924	Patent No. 5496924	383	5	2.8	28	2	US-08-482-142-127	Sequence 127, App
311	5	2.8	14	5	PCT-US91-05177-6	Sequence 6, Appl1	Sequence 6, Appl1	384	5	2.8	28	2	US-08-482-142-128	Sequence 128, App
312	5	2.8	14	6	PCT-US93-07261-9	Sequence 9, Appl1	Sequence 9, Appl1	385	5	2.8	28	2	US-08-482-142-129	Sequence 129, App
313	5	2.8	14	6	5496924-48	Patent No. 5496924	Patent No. 5496924	386	5	2.8	28	2	US-08-488-351A-38	Sequence 38, Appl
314	5	2.8	15	1	US-08-139-862-8	Sequence 8, Appl1	Sequence 8, Appl1	387	5	2.8	28	2	US-08-348-353-34	Sequence 34, Appl
315	5	2.8	15	1	US-08-139-862-14	Sequence 14, Appl	Sequence 14, Appl	388	5	2.8	28	2	US-08-465-965-34	Sequence 34, Appl
316	5	2.8	15	2	US-08-592-646A-53	Sequence 53, Appl	Sequence 53, Appl	389	5	2.8	28	2	US-08-478-572-15	Sequence 15, Appl
317	5	2.8	15	4	US-08-913-362-16	Sequence 16, Appl	Sequence 16, Appl	390	5	2.8	28	2	US-08-478-572-33	Sequence 33, Appl
318	5	2.8	15	4	US-08-913-362-17	Sequence 17, Appl	Sequence 17, Appl	391	5	2.8	28	2		
319	5	2.8	15	4	US-09-165-422-53	Sequence 53, Appl	Sequence 53, Appl	392	5	2.8	28	2		

393	5	2.8	28	2	US-08-478-572-126	Sequence 126, App	466	5	2.8	42	1	US-08-477-383-23	Sequence 23, Appl
394	5	2.8	28	2	US-08-478-572-127	Sequence 127, App	467	5	2.8	42	1	US-08-487-174-23	Sequence 23, Appl
395	5	2.8	28	2	US-08-478-572-128	Sequence 128, App	468	5	2.8	42	1	US-08-480-750-23	Sequence 23, Appl
396	5	2.8	28	2	US-08-478-572-129	Sequence 129, App	469	5	2.8	42	2	US-08-348-353-35	Sequence 35, Appl
397	5	2.8	28	3	US-08-465-966-34	Sequence 34, Appl	470	5	2.8	42	2	US-08-465-965-35	Sequence 35, Appl
398	5	2.8	28	3	US-09-100-414B-37	Sequence 37, Appl	471	5	2.8	42	3	US-08-465-966-35	Sequence 35, Appl
399	5	2.8	28	4	US-09-303-323-37	Sequence 37, Appl	472	5	2.8	44	4	US-09-450-072-76	Sequence 76, Appl
400	5	2.8	28	4	US-08-484-296-15	Sequence 15, Appl	473	5	2.8	44	4	US-09-351-348-76	Sequence 76, Appl
401	5	2.8	28	4	US-08-484-296-33	Sequence 33, Appl	474	5	2.8	45	2	US-08-348-353-6	Sequence 6, Appl
402	5	2.8	28	4	US-08-484-296-126	Sequence 126, App	475	5	2.8	45	2	US-08-348-353-22	Sequence 22, Appl
403	5	2.8	28	4	US-08-484-296-127	Sequence 127, App	476	5	2.8	45	2	US-08-465-965-6	Sequence 6, Appl
404	5	2.8	28	4	US-08-484-296-128	Sequence 128, App	477	5	2.8	45	2	US-08-465-966-22	Sequence 22, Appl
405	5	2.8	28	4	US-08-484-296-129	Sequence 129, App	478	5	2.8	45	3	US-08-465-966-6	Sequence 6, Appl
406	5	2.8	28	2	US-08-482-142-34	Sequence 34, Appl	479	5	2.8	45	3	US-08-465-966-22	Sequence 22, Appl
407	5	2.8	29	2	US-08-482-142-153	Sequence 153, App	480	5	2.8	45	3	US-08-465-966-22	Sequence 22, Appl
408	5	2.8	29	2	US-08-482-142-154	Sequence 154, App	481	5	2.8	45	4	US-09-227-357-166	Sequence 166, App
409	5	2.8	29	2	US-08-478-572-34	Sequence 34, Appl	482	5	2.8	46	1	US-08-446-692-40	Sequence 40, Appl
410	5	2.8	29	2	US-08-478-572-153	Sequence 153, App	483	5	2.8	46	2	US-08-488-351A-40	Sequence 40, Appl
411	5	2.8	29	2	US-08-478-572-154	Sequence 154, App	484	5	2.8	47	4	US-09-227-357-251	Sequence 251, App
412	5	2.8	29	4	US-08-484-296-34	Sequence 34, Appl	485	5	2.8	48	4	US-09-149-476-367	Sequence 367, App
413	5	2.8	29	4	US-08-484-296-153	Sequence 153, App	486	5	2.8	49	4	US-09-004-406C-21	Sequence 21, Appl
414	5	2.8	29	4	US-08-484-296-154	Sequence 154, App	487	5	2.8	51	2	US-08-392-625-30	Sequence 30, Appl
415	5	2.8	30	2	US-08-482-142-152	Sequence 152, App	488	5	2.8	51	2	US-08-466-961A-30	Sequence 30, Appl
416	5	2.8	30	2	US-08-478-572-152	Sequence 152, App	489	5	2.8	51	4	US-09-150-460B-16	Sequence 16, Appl
417	5	2.8	30	4	US-08-484-296-152	Sequence 152, App	490	5	2.8	53	4	US-09-586-563C-6	Sequence 6, Appl
418	5	2.8	30	4	US-09-425-638A-102	Sequence 102, App	491	5	2.8	53	4	US-09-586-562C-6	Sequence 6, Appl
419	5	2.8	30	4	US-09-543-004-102	Sequence 102, App	492	5	2.8	54	4	US-09-605-785-557	Sequence 557, App
420	5	2.8	30	6	5494663-14	Patent No. 5494663	493	5	2.8	59	1	US-08-470-179-11	Sequence 11, Appl
421	5	2.8	31	2	US-08-482-142-121	Sequence 121, App	494	5	2.8	59	2	US-08-637-759B-99	Sequence 99, Appl
422	5	2.8	31	2	US-08-482-142-122	Sequence 122, App	495	5	2.8	59	3	US-08-871-355A-99	Sequence 99, Appl
423	5	2.8	31	2	US-08-482-142-123	Sequence 123, App	496	5	2.8	59	3	US-08-468-011A-23	Sequence 23, Appl
424	5	2.8	31	2	US-08-482-142-151	Sequence 151, App	497	5	2.8	59	3	US-08-513-974B-25	Sequence 25, Appl
425	5	2.8	31	2	US-08-478-572-121	Sequence 121, App	498	5	2.8	59	4	US-08-776-971-20	Sequence 20, Appl
426	5	2.8	31	2	US-08-478-572-122	Sequence 122, App	499	5	2.8	59	4	US-09-236-468A-23	Sequence 23, Appl
427	5	2.8	31	2	US-08-478-572-123	Sequence 123, App	500	5	2.8	59	4	US-09-201-945-99	Sequence 99, Appl
428	5	2.8	31	2	US-08-478-572-123	Sequence 123, App	501	5	2.8	61	3	US-08-078-683A-16	Sequence 16, Appl
429	5	2.8	31	4	US-08-484-296-121	Sequence 151, App	502	5	2.8	61	3	US-08-605-150A-16	Sequence 16, Appl
430	5	2.8	31	4	US-08-484-296-122	Sequence 121, App	503	5	2.8	61	4	US-08-981-392-21	Sequence 21, Appl
431	5	2.8	31	4	US-08-484-296-123	Sequence 122, App	504	5	2.8	61	4	US-09-134-001C-5072	Sequence 5072, App
432	5	2.8	31	4	US-08-484-296-123	Sequence 123, App	505	5	2.8	62	1	US-08-202-047-20	Sequence 20, Appl
433	5	2.8	32	1	US-08-324-301-1	Sequence 151, App	506	5	2.8	62	3	US-08-964-690-20	Sequence 20, Appl
434	5	2.8	32	2	US-08-482-142-124	Sequence 124, App	507	5	2.8	62	4	US-09-134-001C-4525	Sequence 4525, App
435	5	2.8	32	2	US-08-482-142-125	Sequence 125, App	508	5	2.8	63	4	US-09-227-357-213	Sequence 213, App
436	5	2.8	32	2	US-08-482-142-163	Sequence 163, App	509	5	2.8	63	4	US-08-971-089-2	Sequence 2, Appl
437	5	2.8	32	2	US-08-482-142-164	Sequence 164, App	510	5	2.8	66	4	US-09-134-001C-5421	Sequence 5421, App
438	5	2.8	32	2	US-08-478-572-124	Sequence 164, App	511	5	2.8	68	3	US-08-513-974B-311	Sequence 311, App
439	5	2.8	32	2	US-08-478-572-125	Sequence 124, App	512	5	2.8	68	4	US-08-776-971-128	Sequence 128, App
440	5	2.8	32	2	US-08-478-572-125	Sequence 125, App	513	5	2.8	69	4	US-09-149-476-617	Sequence 617, App
441	5	2.8	32	4	US-08-478-572-164	Sequence 164, App	514	5	2.8	70	4	US-09-134-001C-4666	Sequence 4666, App
442	5	2.8	32	4	US-08-484-296-124	Sequence 164, App	515	5	2.8	71	3	US-08-297-395-8	Sequence 8, Appl
443	5	2.8	32	4	US-08-484-296-125	Sequence 125, App	516	5	2.8	72	3	US-09-100-804-19	Sequence 90, Appl
444	5	2.8	32	4	US-08-484-296-163	Sequence 163, App	517	5	2.8	72	3	US-08-348-353-18	Sequence 18, Appl
445	5	2.8	32	4	US-08-484-296-164	Sequence 164, App	518	5	2.8	77	2	US-08-465-965-18	Sequence 18, Appl
446	5	2.8	33	4	US-09-447-125B-26	Sequence 26, Appl	519	5	2.8	77	3	US-08-465-966-18	Sequence 18, Appl
447	5	2.8	36	1	US-08-139-862-6	Sequence 6, Appl	520	5	2.8	77	4	US-09-187-789-25	Sequence 25, Appl
448	5	2.8	36	4	US-09-106-568E-142	Sequence 142, App	521	5	2.8	77	4	US-09-139-600-20	Sequence 20, Appl
449	5	2.8	38	1	US-08-176-500-46	Sequence 46, Appl	522	5	2.8	78	2	US-08-645-193B-16	Sequence 16, Appl
450	5	2.8	38	1	US-08-176-500-123	Sequence 123, App	523	5	2.8	80	2	US-08-691-814B-30	Sequence 30, Appl
451	5	2.8	38	1	US-08-471-052A-46	Sequence 46, Appl	524	5	2.8	82	2	US-08-403-852D-27	Sequence 27, Appl
452	5	2.8	38	1	US-08-471-052A-123	Sequence 123, App	525	5	2.8	82	3	US-08-510-646B-28	Sequence 28, Appl
453	5	2.8	38	1	US-08-189-331-46	Sequence 46, Appl	526	5	2.8	82	4	US-09-231-818-27	Sequence 27, Appl
454	5	2.8	38	1	US-08-189-331-123	Sequence 123, App	527	5	2.8	83	1	US-07-748-783-2	Sequence 2, Appl
455	5	2.8	38	2	US-08-471-939-46	Sequence 46, Appl	528	5	2.8	83	1	US-08-166-818-2	Sequence 2, Appl
456	5	2.8	38	2	US-08-471-939-123	Sequence 123, App	529	5	2.8	84	3	US-09-013-067A-4	Sequence 4, Appl
457	5	2.8	38	2	US-08-471-800-46	Sequence 46, Appl	530	5	2.8	86	1	US-07-869-933-27	Sequence 27, Appl
458	5	2.8	38	2	US-08-471-800-123	Sequence 123, App	531	5	2.8	86	4	US-09-103-663-27	Sequence 27, Appl
459	5	2.8	38	2	US-08-471-068-46	Sequence 46, Appl	532	5	2.8	86	5	PCT-US95-17083-10	Sequence 10, Appl
460	5	2.8	38	2	US-08-471-068-123	Sequence 123, App	533	5	2.8	86	6	US-08-816-155B-47	Sequence 47, Appl
461	5	2.8	40	6	5342615-6	Patent No. 5342615	534	5	2.8	88	3	US-08-894-173-64	Sequence 64, Appl
462	5	2.8	41	6	US-08-640-847C-41	Sequence 41, Appl	535	5	2.8	88	4	US-09-079-587-47	Sequence 47, Appl
463	5	2.8	41	6	US-08-469-260A-539	Sequence 539, App	536	5	2.8	88	4	US-09-398-193-64	Sequence 64, Appl
464	5	2.8	41	6	5196523-6	Patent No. 5196523	537	5	2.8	89	1	US-08-216-276A-27	Sequence 27, Appl
465	5	2.8	42	1	US-08-137-800-23	Sequence 23, Appl	538	5	2.8	91	2	US-08-637-759B-60	Sequence 60, Appl









977	5	2, 8	209	4	US-09-109-100-9.	Sequence 9, Appl
978	5	2, 8	209	4	US-09-109-100-11	Sequence 11, Appl
979	5	2, 8	209	4	US-09-109-100-12	Sequence 12, Appl
980	5	2, 8	209	4	US-09-109-100-13	Sequence 13, Appl
981	5	2, 8	209	4	US-09-109-100-14	Sequence 14, Appl
982	5	2, 8	209	4	US-09-109-100-15	Sequence 15, Appl
983	5	2, 8	209	4	US-09-109-100-16	Sequence 16, Appl
984	5	2, 8	209	4	US-09-109-100-17	Sequence 17, Appl
985	5	2, 8	209	4	US-09-109-100-18	Sequence 18, Appl
986	5	2, 8	209	4	US-08-506-2968-61	Sequence 61, Appl
987	5	2, 8	210	4	US-09-897-537A-2	Sequence 2, Appl
988	5	2, 8	212	1	US-08-792-0198-9	Sequence 9, Appl
989	5	2, 8	212	3	US-08-988-819-9	Sequence 9, Appl
990	5	2, 8	212	4	US-09-016-534-9	Sequence 9, Appl
991	5	2, 8	212	4	US-08-612-973-4	Sequence 4, Appl
992	5	2, 8	212	4	US-08-097-869-7	Sequence 7, Appl
993	5	2, 8	212	4	US-08-795-473B-6	Sequence 6, Appl
994	5	2, 8	212	4	US-08-927-597-4	Sequence 4, Appl
995	5	2, 8	212	4	US-09-320-637-45	Sequence 45, Appl
996	5	2, 8	212	4	US-09-135-941-1	Sequence 1, Appl
997	5	2, 8	212	4	US-09-109-100-10	Sequence 10, Appl
998	5	2, 8	212	4	US-08-702-525-63	Sequence 63, Appl
999	5	2, 8	212	4	US-09-320-371A-27	Sequence 27, Appl
1000	5	2, 8	212	4	US-09-439-856-6	Sequence 6, Appl

## ALIGNMENTS

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Db 390 GATYKFTPNLSVD 402

## RESULT 3

US-08-472-172-4

; Sequence 4, Application US/08472172  
; Patent No. 5985288  
; GENERAL INFORMATION:  
; APPLICANT: Munson, Jr., Robert S  
; APPLICANT: Grass, Susan  
; APPLICANT: Chong, Pele Y Y  
; APPLICANT: Fahim, Raafat  
; APPLICANT: Sia, Charles D Y  
; APPLICANT: McVerry, Patrick  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Outer Membrane Protein p1 and Peptides  
; OF Haemophilus Influenzae Type B  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R6  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/472,172  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/849,411  
; FILING DATE: 07-JUL-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-471  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; TELEX: 063-24567 SIMBAS  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 455 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-472-172-4

Query Match 7.3%; Score 13; DB 2; Length 455;  
Best Local Similarity 100.0%; Pred. No. 2.9e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Caps 0;

. QY 121 GATYKFTPNLSVD 133

|||||

Db 393 GATYKFTPNLSVD 405

## RESULT 4

US-08-472-172-2

; Sequence 2, Application US/08472172  
; Patent No. 5985288  
; GENERAL INFORMATION:  
; APPLICANT: Munson, Jr., Robert S  
; APPLICANT: Grass, Susan  
; APPLICANT: Chong, Pele Y Y  
; APPLICANT: Fahim, Raafat  
; APPLICANT: Sia, Charles D Y

; APPLICANT: McVerry, Patrick  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Outer Membrane Protein p1 and Peptides  
; OF Haemophilus Influenzae Type B  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R6  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/472,172  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/849,411  
; FILING DATE: 07-JUL-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-471  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; TELEX: 063-24567 SIMBAS  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 459 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-472-172-2

Query Match 7.3%; Score 13; DB 2; Length 459;  
Best Local Similarity 100.0%; Pred. No. 3e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 121 GATYKFTPNLSVD 133

|||||

Db 397 GATYKFTPNLSVD 409

## RESULT 5

US-08-475-989-8

; Sequence 8, Application US/08475989  
; Patent No. 5679352  
; GENERAL INFORMATION:  
; APPLICANT: CHONG, Pele  
; APPLICANT: KANDIL, Ali  
; APPLICANT: SIA, Charles  
; APPLICANT: KLEIN, Michel  
; TITLE OF INVENTION: Synthetic Haemophilus Influenzae  
; OF Haemophilus Influenzae  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,989  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/256,839  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA93/00041  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9202219.3  
FILING DATE: 03-FEB-1992  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, MICHAEL I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-505 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 34 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-475-989-8

Query Match 5.6%; Score 10; DB 1; Length 34;  
Best Local Similarity 100.0%; Pred. No. 0.003;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 24 ELSGFHQLTD 33  
Db 2 ELSGFHQLTD 11

RESULT 6  
US-08-475-985-8  
Sequence 8, Application US/08475985  
Patent No. 597349  
GENERAL INFORMATION:  
APPLICANT: CHONG, Pele  
APPLICANT: KANDIL, Ali  
APPLICANT: SIA, Charles  
APPLICANT: KLEIN, Michel  
TITLE OF INVENTION: Synthetic Haemophilus Influenzae  
TITLE OF INVENTION: Conjugate Vaccine  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R6  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,985  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/256,839  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA93/00041  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9202219.3  
FILING DATE: 03-FEB-1992  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, MICHAEL I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-506 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 34 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-475-985-8  
Query Match 5.6%; Score 10; DB 2; Length 34;  
Best Local Similarity 100.0%; Pred. No. 0.003;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 24 ELSGFHQLTD 33  
Db 2 ELSGFHQLTD 11

RESULT 7  
US-08-472-172-20  
Sequence 20, Application US/08472172  
Patent No. 5985288  
GENERAL INFORMATION:  
APPLICANT: Munson, Jr., Robert S  
APPLICANT: Grass, Susan  
APPLICANT: Chong, Pele Y  
APPLICANT: Fahim, Raafat  
APPLICANT: Sia, Charles D Y  
APPLICANT: McVerry, Patrick  
APPLICANT: Klein, Michel  
TITLE OF INVENTION: Outer Membrane Protein P1 and Peptides  
of Haemophilus Influenzae Type B  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R6  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,172  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/849,411  
FILING DATE: 07-JUL-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24973  
REFERENCE/DOCKET NUMBER: 1038-471  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163  
 TELEX: 063-24567 SIMBAS  
 INFORMATION FOR SEQ ID NO: 20:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 34 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-472-172-20

Query Match 5.6%; Score 10; DB 2; Length 34;  
 Best Local Similarity 100.0%; Pred. No. 0.003;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 ELSGFHQLTD 33  
 |||||  
 Db 2 ELSGFHQLTD 11

## RESULT 8

US-08-256-839-8  
 Sequence 8, Application US/08256839  
 Patent No. 6018019

GENERAL INFORMATION:  
 APPLICANT: CHONG, Pele  
 APPLICANT: KANDIL, Ali  
 APPLICANT: SIA, Charles  
 APPLICANT: KLEIN, Michel  
 TITLE OF INVENTION: Synthetic Haemophilus Influenzae  
 TITLE OF INVENTION: Conjugate Vaccine  
 NUMBER OF SEQUENCES: 56  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sim & McBurney  
 STREET: Suite 701, 330 University Avenue  
 CITY: Toronto  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: M5G 1R7

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION NUMBER: US/08/256,839  
 FILING DATE: 03-FEB-1993  
 CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:  
 NAME: STEWART, MICHAEL I.  
 REGISTRATION NUMBER: 24,973  
 REFERENCE/DOCKET NUMBER: 1038-373 MIS:jb  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (416) 595-1155  
 TELEFAX: (416) 595-1163  
 INFORMATION FOR SEQ ID NO: 8:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 34 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-256-839-8

Query Match 5.6%; Score 10; DB 3; Length 34;  
 Best Local Similarity 100.0%; Pred. No. 0.003;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 ELSGFHQLTD 33  
 |||||  
 Db 2 ELSGFHQLTD 11

## RESULT 9

US-08-475-989-14

Sequence 14, Application US/08475989  
 Patent No. 5679352  
 GENERAL INFORMATION:  
 APPLICANT: CHONG, Pele  
 APPLICANT: KANDIL, Ali  
 APPLICANT: SIA, Charles  
 APPLICANT: KLEIN, Michel  
 TITLE OF INVENTION: Synthetic Haemophilus Influenzae  
 TITLE OF INVENTION: Conjugate Vaccine  
 NUMBER OF SEQUENCES: 56  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sim & McBurney  
 STREET: Suite 701, 330 University Avenue  
 CITY: Toronto  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: M5G 1R7

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION NUMBER: US/08/475,989  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:  
 NAME: STEWART, MICHAEL I.  
 REGISTRATION NUMBER: 24,973  
 REFERENCE/DOCKET NUMBER: 1038-505 MIS:vg  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (416) 595-1155  
 TELEFAX: (416) 595-1163  
 INFORMATION FOR SEQ ID NO: 14:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 35 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-475-989-14

Query Match 5.6%; Score 10; DB 1; Length 35;  
 Best Local Similarity 100.0%; Pred. No. 0.0031;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 ANLYGLNLNY 177  
 |||||  
 Db 24 ANLYGLNLNY 33

## RESULT 10

US-08-475-985-14  
 Sequence 14, Application US/08475985  
 Patent No. 5972349

GENERAL INFORMATION:  
 APPLICANT: CHONG, Pele  
 APPLICANT: KANDIL, Ali  
 APPLICANT: SIA, Charles  
 APPLICANT: KLEIN, Michel  
 TITLE OF INVENTION: Synthetic Haemophilus Influenzae  
 TITLE OF INVENTION: Conjugate Vaccine

NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,985  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/256,839  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA93/00041  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9202219.3  
FILING DATE: 03-FEB-1992  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, MICHAEL I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-506 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-475-985-14

Query Match 5.6%; Score 10; DB 2; Length 35;  
Best Local Similarity 100.0%; Pred. No. 0.0031;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 ANLYGLNLNY 177  
|||||  
Db 24 ANLYGLNLNY 33

RESULT 11  
US-08-472-172-26  
Sequence 26, Application US/08472172  
Patent No. 5985288  
GENERAL INFORMATION:  
APPLICANT: Munson, Jr., Robert S  
APPLICANT: Grass, Susan  
APPLICANT: Chong, Pele Y  
APPLICANT: Fahim, Raafat  
APPLICANT: Sia, Charles D Y  
APPLICANT: McVerry, Patrick  
APPLICANT: Klein, Michel  
TITLE OF INVENTION: Outer Membrane Protein p1 and Peptides  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario

COUNTRY: Canada  
ZIP: M5G 1R6  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,172  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/849,411  
FILING DATE: 07-JUL-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24973  
REFERENCE/DOCKET NUMBER: 1038-471  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
TELEX: 063-24567 SIMBAS  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-472-172-26

Query Match 5.6%; Score 10; DB 2; Length 35;  
Best Local Similarity 100.0%; Pred. No. 0.0031;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 ANLYGLNLNY 177  
|||||  
Db 24 ANLYGLNLNY 33

RESULT 12  
US-08-256-839-14  
Sequence 14, Application US/08256839  
Patent No. 6018019  
GENERAL INFORMATION:  
APPLICANT: CHONG, Pele  
APPLICANT: KANDIL, Ali  
APPLICANT: SIA, Charles  
APPLICANT: KLEIN, Michel  
TITLE OF INVENTION: Synthetic Haemophilus Influenzae  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/256,839  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, MICHAEL I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-373 MIS:jb  
TELECOMMUNICATION INFORMATION:



TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-256-839-14

Query Match 5.6%; Score 10; DB 3; Length 35;  
Best Local Similarity 100.0%; Pred. No. 0.0031;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 ANLYGLNLNY 177  
DB 24 ANLYGLNLNY 33  
|||||

## RESULT 13

US-08-475-989-13  
Sequence 13, Application US/08475989  
Patent No. 5679352

GENERAL INFORMATION:  
APPLICANT: CHONG, Pele  
APPLICANT: KANDIL, Ali  
APPLICANT: SIA, Charles  
APPLICANT: KLEIN, Michel  
TITLE OF INVENTION: Synthetic Haemophilus Influenzae  
TITLE OF INVENTION: Conjugate Vaccine  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,989  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/256,839  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA93/00041  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9202219.3  
FILING DATE: 03-FEB-1992  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA93/00041  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9202219.3  
FILING DATE: 03-FEB-1992  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA93/00041  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424

REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-505 MIS:vg  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 38 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-475-989-13

Query Match 5.6%; Score 10; DB 1; Length 38;  
Best Local Similarity 100.0%; Pred. No. 0.0034;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 ANLYGLNLNY 177  
DB 27 ANLYGLNLNY 36  
|||||

## RESULT 14

US-08-475-985-13  
Sequence 13, Application US/08475985  
Patent No. 5972349

GENERAL INFORMATION:  
APPLICANT: CHONG, Pele  
APPLICANT: KANDIL, Ali  
APPLICANT: SIA, Charles  
APPLICANT: KLEIN, Michel  
TITLE OF INVENTION: Synthetic Haemophilus Influenzae  
TITLE OF INVENTION: Conjugate Vaccine  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,985  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/256,839  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA93/00041  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9202219.3  
FILING DATE: 03-FEB-1992  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA93/00041  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424

REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-506 MIS:vg  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 38 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-475-985-13

Query Match 5.6%; Score 10; DB 2; Length 38;  
Best Local Similarity 100.0%; Pred. No. 0.0034;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 ANLYGLNLNY 177  
DB 27 ANLYGLNLNY 36  
|||||

RESULT 15  
US-08-472-172-25  
; Sequence 25, Application US/08472172  
; Patent No. 5985288  
; GENERAL INFORMATION:  
; APPLICANT: Munson, Jr., Robert S  
; APPLICANT: Grass, Susan  
; APPLICANT: Chong, Pele Y Y  
; APPLICANT: Fahim, Raafat  
; APPLICANT: Sia, Charles D Y  
; APPLICANT: McVerry, Patrick  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Outer Membrane Protein P1 and Peptides  
; TITLE OF INVENTION: Of Haemophilus Influenzae Type B  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R6  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/472,172  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/849,411  
; FILING DATE: 07-JUL-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-471  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1163  
; TELEFAX: (416) 595-1163  
; TELEX: 063-24567 SIMBAS  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 38 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-472-172-25

Query Match 5.6%; Score 10; DB 2; Length 38;  
Best Local Similarity 100.0%; Pred. No. 0.0034;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Caps 0;  
Qy 168 ANLYGLNLNY 177  
Db 27 ANLYGLNLNY 36

RESULT 16  
US-08-256-839-13  
; Sequence 13, Application US/08256839  
; Patent No. 6018019  
; GENERAL INFORMATION:  
; APPLICANT: Chong, Pele  
; APPLICANT: KANDIL, Ali  
; APPLICANT: Sia, Charles  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Synthetic Haemophilus Influenzae  
; TITLE OF INVENTION: Conjugate Vaccine  
; NUMBER OF SEQUENCES: 56

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/256,839  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, MICHAEL I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-373 Mis:jfb  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 38 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-256-839-13

Query Match 5.6%; Score 10; DB 3; Length 38;  
Best Local Similarity 100.0%; Pred. No. 0.0034;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Caps 0;  
Qy 168 ANLYGLNLNY 177  
Db 27 ANLYGLNLNY 36

RESULT 17  
US-08-475-989-48  
; Sequence 48, Application US/08475989  
; Patent No. 5679352  
; GENERAL INFORMATION:  
; APPLICANT: CHONG, Pele  
; APPLICANT: KANDIL, Ali  
; APPLICANT: Sia, Charles  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Synthetic Haemophilus Influenzae  
; TITLE OF INVENTION: Conjugate Vaccine  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/475,989  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/256,839  
; FILING DATE: 03-FEB-1993  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/CA93/00041  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA: GB 9202219.3  
FILING DATE: 03-FEB-1992  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, MICHAEL I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-505 MIS:vg  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 48:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 53 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-475-989-48

Query Match 5.6%; Score 10; DB 1; Length 53;  
Best Local Similarity 100.0%; Pred. No. 0.0046;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 ANYGLNLY 177  
Db 42 ANYGLNLY 51

RESULT 18  
US-08-475-985-48  
Sequence 48, Application US/08475985  
Patent No. 5972349  
GENERAL INFORMATION:  
APPLICANT: CHONG, Pele  
APPLICANT: KANDIL, Ali  
APPLICANT: SIA, Charles  
APPLICANT: KLEIN, Michel  
TITLE OF INVENTION: Synthetic Haemophilus Influenzae  
TITLE OF INVENTION: Conjugate Vaccine  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,985  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/256,839  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA93/00041  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9202219.3  
FILING DATE: 03-FEB-1992  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, MICHAEL I.

REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-506 MIS:vg  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 48:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 53 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-475-985-48

Query Match 5.6%; Score 10; DB 2; Length 53;  
Best Local Similarity 100.0%; Pred. No. 0.0046;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 ANYGLNLY 177  
Db 42 ANYGLNLY 51

RESULT 19  
US-08-256-839-48  
Sequence 48, Application US/08256839  
Patent No. 6018019  
GENERAL INFORMATION:  
APPLICANT: CHONG, Pele  
APPLICANT: KANDIL, Ali  
APPLICANT: SIA, Charles  
APPLICANT: KLEIN, Michel  
TITLE OF INVENTION: Synthetic Haemophilus Influenzae  
TITLE OF INVENTION: Conjugate Vaccine  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/256,839  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, MICHAEL I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-373 MIS:jb  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 48:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 53 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-256-839-48

Query Match 5.6%; Score 10; DB 3; Length 53;  
Best Local Similarity 100.0%; Pred. No. 0.0046;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 ANYGLNLY 177  
Db 42 ANYGLNLY 51

```
RESULT 20
US-08-475-989-49
: Sequence 49, Application US/08475989
: Patent No. 5679352
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: KANDIL, All
: APPLICANT: SIA, Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: Synthetic Haemophilus Influenzae
: NUMBER OF SEQUENCES: 56
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: Suite 701, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/475,989
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/256,839
: FILING DATE: 03-FEB-1993
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/CA93/00041
: FILING DATE: 03-FEB-1993
: CLASSIFICATION: 424
: NAME: STEWART, MICHAEL I.
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-505 MIS:Vg
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 49:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 54 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
US-08-475-989-49
```

```
Query Match 5.6%; Score 10; DB 1; Length 54;
Best Local Similarity 100.0%; Pred. No. 0.0047;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Caps 0;
```

```
Qy 168 ANLYGLNLNY 177
Db 43 ANLYGLNLNY 52
```

```
RESULT 21
US-08-475-985-49
: Sequence 49, Application US/08475985
: Patent No. 5972349
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: KANDIL, All
: APPLICANT: SIA, Charles
```

```
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: Synthetic Haemophilus Influenzae
: NUMBER OF SEQUENCES: 56
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: Suite 701, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/475,985
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/256,839
: FILING DATE: 03-FEB-1993
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/CA93/00041
: FILING DATE: 03-FEB-1993
: CLASSIFICATION: 424
: NAME: STEWART, MICHAEL I.
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-506 MIS:Vg
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 49:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 54 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
US-08-475-985-49
```

```
Query Match 5.6%; Score 10; DB 2; Length 54;
Best Local Similarity 100.0%; Pred. No. 0.0047;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Caps 0;
```

```
Qy 168 ANLYGLNLNY 177
Db 43 ANLYGLNLNY 52
```

```
RESULT 22
US-08-256-839-49
: Sequence 49, Application US/08256839
: Patent No. 6018019
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: KANDIL, All
: APPLICANT: SIA, Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: Synthetic Haemophilus Influenzae
: NUMBER OF SEQUENCES: 56
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: Suite 701, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
```

COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/256,839  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, MICHAEL I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-373 MIS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-256-839-49

Query Match 5.6%; Score 10; DB 3; Length 54;  
Best Local Similarity 100.0%; Pred. No. 0.0047;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 ANYGLNLY 177  
DB 43 ANYGLNLY 52

RESULT 23  
US-08-475-989-47  
Sequence 47, Application US/08475989  
Patent No. 5679352  
GENERAL INFORMATION:  
APPLICANT: CHONG, Pele  
APPLICANT: KANDIL, Ali  
APPLICANT: SIA, Charles  
APPLICANT: KLEIN, Michel  
TITLE OF INVENTION: Synthetic Haemophilus Influenzae  
TITLE OF INVENTION: Conjugate Vaccine  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,989  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/256,839  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA93/00041  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9202219.3

FILING DATE: 03-FEB-1992  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, MICHAEL I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-505 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 55 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-475-989-47  
Query Match 5.6%; Score 10; DB 1; Length 55;  
Best Local Similarity 100.0%; Pred. No. 0.0047;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 ANYGLNLY 177  
DB 44 ANYGLNLY 53

RESULT 24  
US-08-475-985-47  
Sequence 47, Application US/08475985  
Patent No. 5972349  
GENERAL INFORMATION:  
APPLICANT: CHONG, Pele  
APPLICANT: KANDIL, Ali  
APPLICANT: SIA, Charles  
APPLICANT: KLEIN, Michel  
TITLE OF INVENTION: Synthetic Haemophilus Influenzae  
TITLE OF INVENTION: Conjugate Vaccine  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,985  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/256,839  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA93/00041  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9202219.3  
FILING DATE: 03-FEB-1992  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, MICHAEL I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-506 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163

## ; INFORMATION FOR SEQ ID NO: 47:

## ; SEQUENCE CHARACTERISTICS:

; LENGTH: 55 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-475-985-47

## Query Match

Best Local Similarity 5.6%; Score 10; DB 2; Length 55;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 ANLYGLNLNY 177

Db 44 ANLYGLNLNY 53

## RESULT 25

US-08-256-839-47

; Sequence 47, Application US/08256839

; Patent No. 6018019

; GENERAL INFORMATION:

; APPLICANT: CHONG, Pele

; APPLICANT: KANDIL, Ali

; APPLICANT: SIA, Charles

; APPLICANT: KLEIN, Michel

; TITLE OF INVENTION: Synthetic Haemophilus Influenzae

; TITLE OF INVENTION: Conjugate Vaccine

; NUMBER OF SEQUENCES: 56

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim &amp; McBurney

; STREET: Suite 701, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/256.839

; FILING DATE:

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: STEWART, MICHAEL I.

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-373 MIS:jb

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1163

; INFORMATION FOR SEQ ID NO: 47:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 55 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-256-839-47

## Query Match

Best Local Similarity 5.6%; Score 10; DB 3; Length 55;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 ANLYGLNLNY 177

Db 44 ANLYGLNLNY 53

## RESULT 26

US-09-347-926-9

; Sequence 9, Application US/09347926

; Patent No. 6440386

## ; GENERAL INFORMATION:

; APPLICANT: LEUNG, SHUI-ON

; TITLE OF INVENTION: STABILIZED RADIOPHOSPHATE-LABELED PROTEINS

; FILE REFERENCE: 018733/0936

; CURRENT APPLICATION NUMBER: US/09/347.926

; CURRENT FILING DATE: 1999-07-06

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 9

; LENGTH: 12

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-347-926-9

## Query Match

Best Local Similarity 3.9%; Score 7; DB 4; Length 12;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 QDGSQYE 63

Db 2 QDGSQYE 8

## RESULT 27

US-09-300-672-7

; Sequence 7, Application US/09300672

; Patent No. 6248937

; GENERAL INFORMATION:

; APPLICANT: Finkelstein, Ruth R.

; APPLICANT: Lynch, Tim

; APPLICANT: Goodman, Howard M.

; APPLICANT: Wang, Ming-Li

; TITLE OF INVENTION: A TRANSCRIPTION FACTOR REGULATING SEED DEVELOPMENT,

; TITLE OF INVENTION: QUALITY AND COLD-TOLERANCE

; FILE REFERENCE: 480.89(HV)

; CURRENT APPLICATION NUMBER: US/09/300.672

; CURRENT FILING DATE: 1999-04-27

; NUMBER OF SEQ ID NOS: 16

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 7

; LENGTH: 59

; TYPE: PRT

; ORGANISM: AP2 domain protein

US-09-300-672-7

## Query Match

Best Local Similarity 3.9%; Score 7; DB 4; Length 59;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 95 LAYDKAA 101

Db 38 LAYDKAA 44

## RESULT 28

US-08-949-603-11

; Sequence 11, Application US/08949603

; Patent No. 5891859

; GENERAL INFORMATION:

; APPLICANT: Michael F. Thomasow and

; APPLICANT: Eric J. Stockinger

; TITLE OF INVENTION: DNA AND ENCODED PROTEIN

; TITLE OF INVENTION: WHICH REGULATES COLD AND

; TITLE OF INVENTION: DEHYDRATION REGULATED GENES

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ian C. McLeod

; STREET: 2190 Commons Parkway

; CITY: Okemos

; STATE: Michigan

; COUNTRY: USA

; ZIP: 48864

; COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS 5.00  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/949,603  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/706,270  
FILING DATE: September 4, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Ian C. McLeod  
REGISTRATION NUMBER: 20,931  
REFERENCE/DOCKET NUMBER: MSU 4.1-384  
TELEPHONE: (517) 347-4100  
TELEFAX: (517) 347-4103  
TELEX: No. 5891859e  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 61  
TYPE: Amino Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: Polypeptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
ORGANISM: Tobacco  
STRAIN:  
INDIVIDUAL ISOLATE: N/A  
DEVELOPMENTAL STAGE: N/A  
HAPLOTYPE: N/A  
TISSUE TYPE: N/A  
CELL TYPE: N/A  
CELL LINE: N/A  
ORGANELLE: N/A  
IMMEDIATE SOURCE: N/A  
POSITION IN GENOME: N/A  
FEATURE:  
NAME/KEY:  
LOCATION:  
IDENTIFICATION METHOD: sequencing  
OTHER INFORMATION: Figure 2D  
PUBLICATION INFORMATION:  
US-08-949-603-11

Query Match 3.9%; Score 7; DB 2; Length 61;  
Best Local Similarity 100.0%; Pred. No. 6.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 LAYDKAA 101  
DB 40 LAYDKAA 46

RESULT 29  
US-08-706-270A-11  
Sequence 11, Application US/08706270A  
Patent No. 5892009  
GENERAL INFORMATION:  
APPLICANT: Michael F. Thomashow and  
APPLICANT: Eric J. Stockinger  
TITLE OF INVENTION: DNA AND ENCODED PROTEIN  
TITLE OF INVENTION: WHICH REGULATES COLD AND  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ian C. McLeod  
STREET: 2190 Commons Parkway

CITY: Okemos  
STATE: Michigan  
COUNTRY: USA  
ZIP: 48864  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS 5.00  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/706,270A  
FILING DATE: September 4, 1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ian C. McLeod  
REGISTRATION NUMBER: 20,931  
REFERENCE/DOCKET NUMBER: MSU 4.1-310  
TELEPHONE: (517) 347-4100  
TELEFAX: (517) 347-4103  
TELEX: No. 5892009e  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 61  
TYPE: Amino Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: Polypeptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
ORGANISM: Tobacco  
STRAIN:  
INDIVIDUAL ISOLATE: N/A  
DEVELOPMENTAL STAGE: N/A  
HAPLOTYPE: N/A  
TISSUE TYPE: N/A  
CELL TYPE: N/A  
CELL LINE: N/A  
ORGANELLE: N/A  
IMMEDIATE SOURCE: N/A  
POSITION IN GENOME: N/A  
FEATURE:  
NAME/KEY:  
LOCATION:  
IDENTIFICATION METHOD: sequencing  
OTHER INFORMATION: Figure 2D  
PUBLICATION INFORMATION:  
US-08-706-270A-11

Query Match 3.9%; Score 7; DB 2; Length 61;  
Best Local Similarity 100.0%; Pred. No. 6.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 LAYDKAA 101  
DB 40 LAYDKAA 46

RESULT 30  
US-08-949-580-11  
Sequence 11, Application US/08949580  
Patent No. 5929305  
GENERAL INFORMATION:  
APPLICANT: Michael F. Thomashow and  
APPLICANT: Eric J. Stockinger  
TITLE OF INVENTION: DNA AND ENCODED PROTEIN  
TITLE OF INVENTION: WHICH REGULATES COLD AND

;; TITLE OF INVENTION: DEHYDRATION REGULATED GENES  
;; NUMBER OF SEQUENCES: 11  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Ian C. McLeod  
;; STREET: 2190 Commons Parkway  
;; CITY: Okemos  
;; STATE: Michigan  
;; COUNTRY: USA  
;; ZIP: 48864  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb  
;; MEDIUM TYPE: storage  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: MS-DOS 5.00  
;; SOFTWARE: Wordperfect 5.1  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/949,580  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; REGISTRATION NUMBER: 20,931  
;; REFERENCE/DOCKET NUMBER: MSU 4.1-383  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (517) 347-4100  
;; TELEFAX: (517) 347-4103  
;; TELEX: No. 5929305e  
;; INFORMATION FOR SEQ ID NO: 11:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 61  
;; TYPE: Amino Acid  
;; STRANDEDNESS: Single  
;; TOPOLOGY: Linear  
;; MOLECULE TYPE: Polypeptide  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; FRAGMENT TYPE:  
;; ORIGINAL SOURCE:  
;; ORGANISM: Tobacco  
;; STRAIN:  
;; INDIVIDUAL ISOLATE: N/A  
;; DEVELOPMENTAL STAGE: N/A  
;; HAPLOTYPE: N/A  
;; TISSUE TYPE: N/A  
;; CELL TYPE: N/A  
;; CELL LINE: N/A  
;; ORGANELLE: N/A  
;; IMMEDIATE SOURCE: N/A  
;; POSITION IN GENOME: N/A  
;; FEATURE:  
;; NAME/KEY:  
;; LOCATION:  
;; IDENTIFICATION METHOD: sequencing  
;; OTHER INFORMATION: Figure 2D  
;; PUBLICATION INFORMATION:  
US-08-949-580-11

Query Match 3.9%; Score 7; DB 2; Length 61;  
Best Local Similarity 100.0%; Pred. No. 6.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 95 LAYDKAA 101  
|||||  
Db 40 LAYDKAA 46

RESULT 31  
US-08-950-172A-11  
; Sequence 11, Application US/08950172A  
; Patent No. 5965705

;; GENERAL INFORMATION:  
;; APPLICANT: Michael F. Thomashow and  
;; APPLICANT: Eric J. Stockinger  
;; TITLE OF INVENTION: DNA AND ENCODED PROTEIN  
;; TITLE OF INVENTION: WHICH REGULATES COLD AND  
;; TITLE OF INVENTION: DEHYDRATION REGULATED GENES  
;; NUMBER OF SEQUENCES: 11  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Ian C. McLeod  
;; STREET: 2190 Commons Parkway  
;; CITY: Okemos  
;; STATE: Michigan  
;; COUNTRY: USA  
;; ZIP: 48864  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb  
;; MEDIUM TYPE: storage  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: MS-DOS 5.00  
;; SOFTWARE: Wordperfect 5.1  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/950,172A  
;; FILING DATE: 10/14/1997  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/949,580  
;; FILING DATE: October 14, 1997  
;; APPLICATION NUMBER: 08/706,270  
;; FILING DATE: September 4, 1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Ian C. McLeod  
;; REGISTRATION NUMBER: 20,931  
;; REFERENCE/DOCKET NUMBER: MSU 4.1-383  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (517) 347-4100  
;; TELEFAX: (517) 347-4103  
;; TELEX: No. 5965705e  
;; INFORMATION FOR SEQ ID NO: 11:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 61  
;; TYPE: Amino Acid  
;; STRANDEDNESS: Single  
;; TOPOLOGY: Linear  
;; MOLECULE TYPE: Polypeptide  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; FRAGMENT TYPE:  
;; ORIGINAL SOURCE:  
;; ORGANISM: Tobacco  
;; STRAIN:  
;; INDIVIDUAL ISOLATE: N/A  
;; DEVELOPMENTAL STAGE: N/A  
;; HAPLOTYPE: N/A  
;; TISSUE TYPE: N/A  
;; CELL TYPE: N/A  
;; CELL LINE: N/A  
;; ORGANELLE: N/A  
;; IMMEDIATE SOURCE: N/A  
;; POSITION IN GENOME: N/A  
;; FEATURE:  
;; NAME/KEY:  
;; LOCATION:  
;; IDENTIFICATION METHOD: sequencing  
;; OTHER INFORMATION: Figure 2D  
;; PUBLICATION INFORMATION:  
US-08-950-172A-11

Query Match 3.9%; Score 7; DB 2; Length 61;  
Best Local Similarity 100.0%; Pred. No. 6.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 95 LAYDKAA 101  
|||||



Db 40 LAYDKAA 46

## RESULT 32

US-09-198-119C-11  
; Sequence 11, Application US/09198119C

; Patent No. 6417428

; GENERAL INFORMATION:

; APPLICANT: Thomashow, Michael

; APPLICANT: Stockinger, Eric

; APPLICANT: Jaglo-Ortosen, Kirsten

; APPLICANT: Gilmour, Sarah

; APPLICANT: Zarka, Daniel

; APPLICANT: Jiang, Cai-zhong

; TITLE OF INVENTION: Plant Having Altered Environmental Stress Tolerance

; FILE REFERENCE: 19117.713 Seq List

; CURRENT APPLICATION NUMBER: US/09/198,119C

; CURRENT FILING DATE: 1998-11-23

; PRIOR APPLICATION NUMBER: US 08/706,270

; PRIOR FILING DATE: 1998-09-04

; PRIOR APPLICATION NUMBER: US 09/018,233

; PRIOR FILING DATE: 1998-02-03

; PRIOR APPLICATION NUMBER: US 09/017,816

; PRIOR FILING DATE: 1998-02-03

; PRIOR APPLICATION NUMBER: US 09/018,235

; PRIOR FILING DATE: 1998-02-03

; PRIOR APPLICATION NUMBER: US 09/017,575

; PRIOR FILING DATE: 1998-02-03

; PRIOR APPLICATION NUMBER: US 09/018,227

; PRIOR FILING DATE: 1998-02-03

; PRIOR APPLICATION NUMBER: US 09/018,234

; PRIOR FILING DATE: 1998-02-03

; NUMBER OF SEQ ID NOS: 95

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 11

; LENGTH: 61

; TYPE: PRT

; ORGANISM: Nicotiana tabacum

US-09-198-119C-11

Query Match 3.9%: Score 7; DB 4; Length 61;

Best Local Similarity 100.0%: Pred. No. 6.2;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 LAYDKAA 101

|||||

Db 40 LAYDKAA 46

## RESULT 33

US-08-912-272-28

; Sequence 28, Application US/08912272

; Patent No. 6093874

; GENERAL INFORMATION:

; APPLICANT: Jofuku, K. Diane

; APPLICANT: Okamuro, Jack K.

; TITLE OF INVENTION: Methods for Improving Seeds

; NUMBER OF SEQUENCES: 103

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/912,272

; FILING DATE: 15-AUG-1997

CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/879,827  
FILING DATE: 20-JUN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/700,152  
FILING DATE: 20-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 023070-067220US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 68 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..68  
OTHER INFORMATION: /note= "EREBP-like Arabidopsis RAP2.4  
OTHER INFORMATION: AP2 domain"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 35..50  
OTHER INFORMATION: /note= "putative RAP2.4 amphipathic  
OTHER INFORMATION: alpha-helix"  
US-08-912-272-28

Query Match

3.9%: Score 7; DB 3; Length 68;

Best Local Similarity 100.0%: Pred. No. 6.9;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 LAYDKAA 101

|||||

Db 41 LAYDKAA 47

## RESULT 34

US-09-026-039-28

; Sequence 28, Application US/09026039

; Patent No. 6329567

; GENERAL INFORMATION:

; APPLICANT: Jofuku, K. Diane

; APPLICANT: Okamuro, Jack K.

; TITLE OF INVENTION: Methods for Improving Seeds

; NUMBER OF SEQUENCES: 103

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/026,039

; FILING DATE: 19-FEB-1998

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/912,272

; FILING DATE: 15-AUG-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/879,827

; FILING DATE: 20-JUN-1997

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/700,152  
;; FILING DATE: 20-AUG-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Bastian, Kevin L.  
;; REGISTRATION NUMBER: 34,774  
;; REFERENCE/DOCKET NUMBER: 023070-067230US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 576-0200  
;; TELEFAX: (415) 576-0300  
;; INFORMATION FOR SEQ ID NO: 28:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 68 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FEATURE:  
;; NAME/KEY: Peptide  
;; LOCATION: 1..68  
;; OTHER INFORMATION: /note= "EREBP-like Arabidopsis RAP2.4  
;; OTHER INFORMATION: AP2 domain"  
;; FEATURE:  
;; NAME/KEY: Region  
;; LOCATION: 35..50  
;; OTHER INFORMATION: /note= "putative RAP2.4 amphipathic  
;; OTHER INFORMATION: alpha-helix"  
;; US-09-026-039-28  
;;  
Query Match 3.9%; Score 7; DB 4; Length 68;  
Best Local Similarity 100.0%; Pred. No. 6.9;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 95 LAYDKAA 101  
Db 41 LAYDKAA 47  
|||||  
RESULT 35  
US-08-912-272-18  
; Sequence 18, Application US/08912272  
; Patent No. 6093874  
; GENERAL INFORMATION:  
; APPLICANT: Jofuku, K. Diane  
; APPLICANT: Okamoto, Jack K.  
; TITLE OF INVENTION: Methods for Improving Seeds  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 15-AUG-1997  
; APPLICATION NUMBER: US/08/912,272  
; FILING DATE: 15-AUG-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/700,152  
; FILING DATE: 20-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774  
; REFERENCE/DOCKET NUMBER: 023070-067220US

;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 576-0200  
;; TELEFAX: (415) 576-0300  
;; INFORMATION FOR SEQ ID NO: 18:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 69 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FEATURE:  
;; NAME/KEY: Peptide  
;; LOCATION: 1..69  
;; OTHER INFORMATION: /note= "AP2 domain within tobacco  
;; OTHER INFORMATION: EREBP-2"  
;; FEATURE:  
;; NAME/KEY: Region  
;; LOCATION: 35..51  
;; OTHER INFORMATION: /note= "putative EREBP-2 amphipathic  
;; OTHER INFORMATION: alpha-helix"  
;; US-08-912-272-18  
;;  
Query Match 3.9%; Score 7; DB 3; Length 69;  
Best Local Similarity 100.0%; Pred. No. 7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 95 LAYDKAA 101  
Db 42 LAYDKAA 48  
|||||  
RESULT 36  
US-09-026-039-18  
; Sequence 18, Application US/09026039  
; Patent No. 6325567  
; GENERAL INFORMATION:  
; APPLICANT: Jofuku, K. Diane  
; APPLICANT: Okamoto, Jack K.  
; TITLE OF INVENTION: Methods for Improving Seeds  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 19-FEB-1998  
; APPLICATION NUMBER: US/09/026,039  
; FILING DATE: 19-FEB-1998  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/912,272  
; FILING DATE: 15-AUG-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/879,827  
; FILING DATE: 20-JUN-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/700,152  
; FILING DATE: 20-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774  
; REFERENCE/DOCKET NUMBER: 023070-067230US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 576-0200  
;; TELEFAX: (415) 576-0300  
;; INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:  
LENGTH: 69 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..69  
OTHER INFORMATION: /note= "AP2 domain within tobacco  
OTHER INFORMATION: EREBP-2"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 35..51  
OTHER INFORMATION: /note= "putative EREBP-2 amphipathic  
OTHER INFORMATION: alpha-helix"  
US-09-026-039-18

Query Match 3.9%; Score 7; DB 4; Length 69;  
Best Local Similarity 100.0%; Pred. No. 7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 95 LAYDKAA 101  
Db 42 LAYDKAA 48

RESULT 37  
US-08-726-136-28  
; Sequence 28, Application US/08726136  
; Patent No. 5811286  
; GENERAL INFORMATION:  
; APPLICANT: ROBERT D. FALLON  
; APPLICANT: MARK S. PAYNE  
; APPLICANT: MARK J. NELSON  
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS ENCODING  
; TITLE OF INVENTION: STEREOSPECIFIC NITRILE HYDRATASE AND AMIDASE ENZYMES AND  
; TITLE OF INVENTION: RECOMBINANT ORGANISMS EXPRESSING THOSE ENZYMES USEFUL FOR  
; TITLE OF INVENTION: THE PRODUCTION OF CHIRAL AMIDES AND ACIDS  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 19898  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MICROSOFT WINDOWS 3.1  
; SOFTWARE: MICROSOFT WORD 2.0C  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/726,136  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/004914  
; FILING DATE: OCTOBER 6, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FLOYD, LINDA A.  
; REGISTRATION NUMBER: 33,692  
; REFERENCE/DOCKET NUMBER: CR-9677  
; TELEPHONE: 302-892-8112  
; TELEFAX: 302-773-0164  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 466 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein

; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; STRAIN: AMIDASE  
US-08-726-136-28

Query Match 3.9%; Score 7; DB 2; Length 466;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 LRAGLAY 97  
Db 443 LRAGLAY 449

RESULT 38  
US-09-103-434-28  
; Sequence 28, Application US/09103434  
; Patent No. 6133421  
; GENERAL INFORMATION:  
; APPLICANT: ROBERT D. FALLON  
; APPLICANT: MARK S. PAYNE  
; APPLICANT: MARK J. NELSON  
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS ENCODING  
; TITLE OF INVENTION: STEREOSPECIFIC NITRILE HYDRATASE AND AMIDASE ENZYMES AND  
; TITLE OF INVENTION: RECOMBINANT ORGANISMS EXPRESSING THOSE ENZYMES USEFUL FOR  
; TITLE OF INVENTION: THE PRODUCTION OF CHIRAL AMIDES AND ACIDS  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 19898  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MICROSOFT WINDOWS 3.1  
; SOFTWARE: MICROSOFT WORD 2.0C  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/103,434  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/726,136  
; FILING DATE:  
; APPLICATION NUMBER: 60/004914  
; FILING DATE: OCTOBER 6, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FLOYD, LINDA A.  
; REGISTRATION NUMBER: 33,692  
; REFERENCE/DOCKET NUMBER: CR-9677  
; TELEPHONE: 302-892-8112  
; TELEFAX: 302-773-0164  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 466 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; STRAIN: AMIDASE  
US-09-103-434-28

Query Match 3.9%; Score 7; DB 4; Length 466;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 Lraglay 97  
| | | | |  
Db 443 Lraglay 449

## RESULT 39

US-09-687-594-28

; Sequence 28, Application US/09687594

; Patent No. 6251650

; GENERAL INFORMATION:

; APPLICANT: ROBERT D. FALLON

; APPLICANT: MARK S. PAYNE

; APPLICANT: MARK J. NELSON

; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS ENCODING

; TITLE OF INVENTION: STEROSPECIFIC NITRILE HYDRATASE AND AMIDASE ENZYMES AND

; TITLE OF INVENTION: RECOMBINANT ORGANISMS EXPRESSING THOSE ENZYMES USEFUL FOR

; TITLE OF INVENTION: THE PRODUCTION OF CHIRAL AMIDES AND ACIDS

; NUMBER OF SEQUENCES: 28

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY

; STREET: 1007 MARKET STREET

; CITY: WILMINGTON

; STATE: DELAWARE

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 19898

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: MICROSOFT WINDOWS 3.1

; SOFTWARE: MICROSOFT WORD 2.0C

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/687,594

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/726,136

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: FLOYD, LINDA A.

; REGISTRATION NUMBER: 33,692

; REFERENCE/DOCKET NUMBER: CR-9677

; TELEPHONE: 302-892-8112

; TELEFAX: 302-773-0164

; INFORMATION FOR SEQ ID NO: 28:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 466 amino acids

; TYPE: amino acid

; STRANDEDNESS: unknown

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; STRAIN: AMIDASE

; US-09-687-594-28

## Query Match

3.9%; Score 7; DB 4; Length 466;

Best Local Similarity 100.0%; Pred. No. 42;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 Lraglay 97  
| | | | |  
Db 443 Lraglay 449

## RESULT 40

US-08-056-200-60

; Sequence 60, Application US/08056200

; Patent No. 5616500

; GENERAL INFORMATION:

; APPLICANT: Steinert, Peter M.

; APPLICANT: Lee, Seung-Chul

; APPLICANT: Kim, In-Gyu  
; APPLICANT: Chung, Soo-Il  
; APPLICANT: Park, Sang-Chul  
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
; TITLE OF INVENTION: Methods of Using Same  
; NUMBER OF SEQUENCES: 117  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive, Sixteenth Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/056,200  
; FILING DATE: 30-APR-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pedrick, Michael F.  
; REGISTRATION NUMBER: 36,799  
; REFERENCE/DOCKET NUMBER: NIH054.001A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (714) 760-0404  
; TELEFAX: (714) 760-9502  
; INFORMATION FOR SEQ ID NO: 60:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; US-08-056-200-60

## Query Match

3.4%; Score 6; DB 1; Length 15;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 NLSVDV 134  
| | | | |  
Db 7 NLSVDV 12

Search completed: May 12, 2003, 09:49:11  
Job time : 30 secs

Result No.	Score	Query Match	Length	DB ID	Description
1	256	26.8	77	12	H.influenzae strai
2	151	15.8	459	21	Neisseria gonorrhoe
3	144.5	15.1	442	21	Neisseria meningit
4	144.5	15.1	466	21	Neisseria meningit
5	143.5	15.0	466	21	Neisseria meningit
6	142.5	14.9	464	21	Neisseria meningit
7	133.5	13.9	456	23	Neisseria meningit
8	129	13.5	35	14	Tmox amino acid se
9	111	11.6	33	12	Hib OMP Pl peptide
10	111	11.6	33	14	HibP1-10(339-370)
					Hib OMP Pl peptide

PT vaccine against infections, esp. in infants and for diagnosis

PS Disclosure; Fig 5; 33pp; English.

CC Plasmid pRSM793 contains only the 3' portion of the p1 gene. The  
CC plasmid is derived from pRSM188 which contains the full-length p1  
CC sequence. This cro-lacZ-ompP1 fusion protein is produced by pRSM793  
CC and is recognised by rabbit and guinea pig p1-specific antisera in  
CC immunoblot analyses.  
CC See also AAR12446-R12455.

XX Sequence 77 AA;

SO Query Match 26.8%; Score 256; DB 12; Length 77;  
Best Local Similarity 64.9%; Pred. No. 2.5e-20;  
Matches 50; Conservative 10; Mismatches 13; Indels 4; Gaps 1;

Qy 107 SASIPDTRMYSGATYKFTPNLSVDVGFALHCKKKHFEVETQNIKG-LLLVVEADY 162  
Db 1 SAAIPDTRMYSGATYKFTPNLSVDVGFALHCKKKHFEVETQNIKG-LLLVVEADY 162

Qy 163 TTKATANLYGLNLYRF 179

Db 61 TSQAHANLYGLNLYSF 77

RESULT 2

ID AAY75757 standard; Protein; 459 AA.

XX AAY75757;

XX 21-MAR-2000 (first entry)

DE Neisseria gonorrhoeae ORF 989 protein sequence SEQ ID NO:2986.

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;  
KW antibacterial; gene therapy.

OS Neisseria gonorrhoeae.

PN W09957280-A2.

PD 11-NOV-1999.

PF 30-APR-1999; 99WO-US09346.

PR 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094859.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0099062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

XX (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;

PI Petersen J, Piazza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

PI Tettelin H, Venter JC;

XX WPI; 2000-062150/05.

DR N-PSDB; AA254519.

XX Novel Neisserial polypeptides predicted to be useful antigens for

PT vaccines and diagnostics

XX Claim 2; Page 1404; 1453pp; English.

XX AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941

CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides  
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
CC PCR primers used in the exemplification of the present invention. The  
CC polypeptides, the polynucleotides, antibodies and compositions of  
CC the invention can be used as vaccines, as diagnostic reagents, and as  
CC immunogenic compositions. The polypeptides can be used in the  
CC manufacture of medicaments for treating or preventing infection due to  
CC *Neisserial* bacteria (e.g. meningitis and septicaemia), to detect the  
CC presence of *Neisseria* bacteria, or to raise antibodies. They may also  
CC be used to screen for agonists or antagonists, which may themselves  
CC have use as antibacterial agents. The polynucleotides of the invention  
CC may also be used in gene therapy protocols.

XX Sequence 459 AA;

SO Query Match 15.8%; Score 151; DB 21; Length 459;  
Best Local Similarity 22.3%; Pred. No. 9.3e-08;  
Matches 43; Conservative 36; Mismatches 74; Indels 40; Gaps 5;

Qy 12 KGSITLKLPAWELSGFHLQTDMAIHYSKYKTEWSRFK-----ELRGKYQDGSVGEAF 65

Db 281 KASVKIVTPESLSVHGMVKVSDKADLFGDVTWTRHSRENKAELFEKEKNTANGKSDRT 340

Qy 66 TKKEEYKDNSRFAITGTYSLNDALTLRAGLAYDK--AASKTHLSASIPDTRMYSGAT 123

Db 341 TITPNWRNTYKVGGLGGYQISEPLQLRVGIAFDKPPVRNADYRMNSLPLDGNRIWFSAGMK 400

Qy 124 YKFTPNLSVDVGFALHCKKKHFEVETQNIKGLLVVEADYTTKATAN----- 169

Db 401 YHICKNHVVDAAYTHI-----HINDT-----SYRTAKASGNDVDKSGASCAREFN 445

Qy 170 ---LYGLNLYRF 179

Db 446 HADIIGLYTYKF 458

RESULT 3

ID AAB01259 standard; Protein; 442 AA.

XX AAB01259;

XX 27-OCT-2000 (first entry)

DE *Neisseria meningitidis* BASB044 protein #2.

KW Meningitis; microbial disease; upper respiratory tract infection;

KW bacteraemia; invasive bacterial disease; BASB044 protein.

OS *Neisseria meningitidis*.

XX W0200034482-A2.

XX 15-JUN-2000.

XX 07-DEC-1999; 99WO-IB02014.

XX 08-DEC-1998; 98GB-0026979.

XX 08-DEC-1998; 98GB-0026980.

XX 17-DEC-1998; 98GB-0028015.

XX 05-JAN-1999; 99GB-0000090.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Ruelle J, Verlant VGCL;

XX WPI; 2000-423427/36.

DR N-PSDB; AAA49619.

XX Novel BASB041, 43, 44 and 48 polypeptides of *Neisseria meningitidis*  
PT useful for diagnostic, prophylactic and therapeutic purposes against  
PT microbial diseases comprise a specific amino acid sequence

PS Claim 52; Page 166-167; 171pp; English.

XX The present sequence is the BASB044 protein from the *Neisseria*  
 CC meningitidis strain H44/76. The protein and its gene can be used in  
 CC the prevention and treatment of microbial disease such as bacteraemia,  
 CC meningitis and upper respiratory tract infections. They are particularly  
 CC useful for treating bacterial diseases. They can also be used for  
 CC diagnosing these diseases. The sequence shows strong homology to  
 CC the BASB044 protein from *N. meningitidis* strain ATCC 13090.

XX Sequence 442 AA;

Query Match 15.1%; Score 144.5; DB 21; Length 442;

Best Local Similarity 23.4%; Pred. No. 4.5e-07;

Matches 45; Conservative 38; Mismatches 70; Indels 39; Gaps 7;

QY 12 KGSLLTLKLPAYWELSGFHOLTDQWATHYSKYKTEWSRF-----KE---LRGKYODGS 60  
 Db 265 KARVKIVTPESLSVHGMKYKSDKADLFGDVTWTRHSRFDKAEVLPEKEKTVVYKKG----- 319

QY -61 GYEAFTKKEEYKDNSRFAIGTTYSNLDAITLRAGLAYDKAASKT--HLSASIPDTRMMY 118  
 Db 320 -SDRTTITPNWRNTYKVGFGSGYQISEPLQLRAGIAFDKSPVRNADYRMNSLPDGNRIWF 378

QY 119 SIGATYKFTPNLSVDVGFHAHLRGKKKHVFETO-----NKGLLLVLEADYTTKAT 167  
 Db 379 SAGMKYHIGKNHVDAAVYTHI-----HINDTSYRTAKASGNDVDVSKG-----ASSAREKHNH 429

QY 168 ANLYGLNLNRYF 179

Db 430 ADIIGLQYTYKF 441

RESULT 4

AAV75758

ID AAV75758 standard; Protein; 466 AA.

XX AC AAV75758;

XX DT 21-MAR-2000 (first entry)

XX DE *Neisseria meningitidis* ORF 989 protein sequence SEQ ID NO:2988.

XX KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;  
 KW antibacterial; gene therapy.

XX OS *Neisseria meningitidis*.

XX PN WO9957280-A2.

XX PD 11-NOV-1999.

XX PF 30-APR-1999; 99WO-US09346.

XX PR 01-MAY-1998; 98US-0083758.

XX PR 31-JUL-1998; 98US-0094869.

XX PR 02-SEP-1998; 98US-0098994.

XX PR 02-SEP-1998; 98US-0099062.

XX PR 09-OCT-1998; 98US-0103749.

XX PR 09-OCT-1998; 98US-0103794.

XX PR 09-OCT-1998; 98US-0103796.

XX PR 25-FEB-1999; 99US-0121528.

XX PA (CHIR ) CHIRON CORP.

XX PA (GENO-) INST GENOMIC RES.

XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Masiagnani V, Mora M;  
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;

XX WPI: 2000-062150/05.

DR N-PSDB; AAZ54520.

XX Novel *Neisseria* polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics  
 XX Claim 2; Page 1404-1405; 1453pp; English.

XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941  
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides  
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC *Neisseria* bacteria (e.g. meningitis and septicaemia), to detect the  
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.

XX Sequence 466 AA;

Query Match 15.1%; Score 144.5; DB 21; Length 466;

Best Local Similarity 23.4%; Pred. No. 4.9e-07;

Matches 45; Conservative 38; Mismatches 70; Indels 39; Gaps 7;

QY 12 KGSLLTLKLPAYWELSGFHOLTDQWATHYSKYKTEWSRF-----KE---LRGKYODGS 60

Db 289 KARVKIVTPESLSVHGMKYKSDKADLFGDVTWTRHSRFDKAEVLPEKEKTVVYKKG----- 343

QY 61 GYEAFTKKEEYKDNSRFAIGTTYSNLDAITLRAGLAYDKAASKT--HLSASIPDTRMMY 118

Db 344 -SDRTTITPNWRNTYKVGFGSGYQISEPLQLRAGIAFDKSPVRNADYRMNSLPDGNRIWF 402

QY 119 SIGATYKFTPNLSVDVGFHAHLRGKKKHVFETO-----NKGLLLVLEADYTTKAT 167

Db 403 SAGMKYHIGKNHVDAAVYTHI-----HINDTSYRTAKASGNDVDVSKG-----ASSAREKHNH 453

QY 168 ANLYGLNLNRYF 179

Db 454 ADIIGLQYTYKF 465

RESULT 5

AAAB01258

ID AAB01258 standard; Protein; 466 AA.

XX AC AAB01258;

XX DT 27-OCT-2000 (first entry)

XX DE *Neisseria meningitidis* BASB044 protein #1.

XX KW Meningitis; microbial disease; upper respiratory tract infection;  
 KW bacteraemia; invasive bacterial disease; BASB044 protein.

XX OS *Neisseria meningitidis*.

XX FH Key Location/Qualifiers

FT Peptide 1..24

FT Protein /label= signal\_peptide

FT /label= 25..466

FT /label= mature\_BASB044

XX WO200034482-A2.

XX PD 15-JUN-2000.

XX PF 07-DEC-1999; 99WO-IB02014.

XX PR 08-DEC-1998; 98GB-0026979.

XX PR 08-DEC-1998; 98GB-0026980.

XX PR 17-DEC-1998; 98GB-0028015.

```

PR 05-JAN-1999; 99GB-0000090.
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PI Ruelle J, Verlant VGCL;
XX
XX WPI; 2000-423427/36.
DR N-PSDB; AAA49618.
XX
XX Novel BASB041, 43, 44 and 48 polypeptides of Neisseria meningitidis
PT useful for diagnostic, prophylactic and therapeutic purposes against
PT microbial diseases comprise a specific amino acid sequence
XX
XX Claim 52: Page 163-165; 171pp: English.
XX
XX The present sequence is the BASB044 protein from the Neisseria
CC meningitidis strain ATCC 13090. The protein and its gene can be used in
CC the prevention and treatment of microbial disease such as bacteraemia,
CC meningitis and upper respiratory tract infections. They are particularly
CC useful for treating bacterial diseases. They can also be used for
CC diagnosing these diseases. The protein sequence shows significant
CC homology to the long chain fatty acid transport protein FadL of E. coli.
XX
XX Sequence 466 AA;
XX
XX Query Match 15.0%; Score 143.5; DB 21; Length 466;
XX Best Local Similarity 23.4%; Pred. No. 6.3e-07;
XX Matches 45; Conservative 38; Mismatches 70; Indels 39; Gaps 7;
XX
QY 12 KGSLLKLPAYWELSGFHOLTDOVAIHYSKYKTEWSRF-----KE---LRGKYODGS 60
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
289 KARVKIVTPESLVHGMKYKSDKADLFQDVTWTRHSRFDKAELVFEKEKTVVKGK----- 343
QY 61 GYEAFKKEEYKDNSRFAIGTYSNLDAITLRAGLAYDKAAASKT--HLSASIPDTRMWY 118
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
344 -SDRTTTPNWRNTYKVGFGSGYQISEPLQLRAGIAFDKSPVRNADYRMNSLPDGNRIWF 402
QY 119 SIGATYKFTPNLSVDVGFALHGRGKKHVFETO-----NIKGLLLVEADYTTTKAT 167
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
403 SAGMKYHIGKNHVVDAAAYTHI-----HINDTYSRTAKASGNDVDSKG-----ASSAREKNH 453
QY 168 ANLYGLNLNRYF 179
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
454 ADIIGLQYTYKF 465
XX
XX RESULT 6
XX AAY75759
XX ID AAY75759 standard; Protein; 464 AA.
XX AC AAY75759;
XX
XX 21-MAR-2000 (first entry)
XX
XX Neisseria meningitidis ORF 989 protein sequence SEQ ID NO:2990.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX
XX Neisseria meningitidis.
XX
XX WO9957280-A2.
XX
XX 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 03-OCT-1998; 98US-0103749.
XX
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 98US-0121528.
XX
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Piazza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX WPI; 2000-062150/05.
DR N-PSDB; AA254521.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics
XX
XX Claim 2: Page 1406; 1453pp: English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisseria bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
XX Sequence 464 AA;
XX
XX Query Match 14.9%; Score 142.5; DB 21; Length 464;
XX Best Local Similarity 23.4%; Pred. No. 8e-07;
XX Matches 45; Conservative 37; Mismatches 71; Indels 39; Gaps 7;
XX
QY 12 KGSLLKLPAYWELSGFHOLTDOVAIHYSKYKTEWSRF-----KE---LRGKYODGS 60
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
287 KARVKIVTPESLVHGMKYKSDKADLFQDVTWTRHSRFDKAELVFEKEKTVVNGK----- 341
QY 61 GYEAFKKEEYKDNSRFAIGTYSNLDAITLRAGLAYDKAAASKT--HLSASIPDTRMWY 118
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
342 -SDRTTTPNWRNTYKVGFGSGYQISEPLQLRAGIAFDKSPVRNADYRMNSLPDGNRIWF 400
QY 119 SIGATYKFTPNLSVDVGFALHGRGKKHVFETO-----NIKGLLLVEADYTTTKAT 167
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
401 SAGMKYHIGKNHVVDAAAYTHI-----HINDTYSRTAKASGNDVDSKG-----ASSAREKNH 451
QY 168 ANLYGLNLNRYF 179
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
452 ADIIGLQYTYKF 463
XX
XX RESULT 7
XX ABB08386
XX ID ABB08386 standard; Protein; 456 AA.
XX AC ABB08386;
XX
XX 18-JUN-2002 (first entry)
XX
XX Tmox amino acid sequence.
XX
XX PHBA; para-Hydrobenzoate; liquid crystal polymer; LCP;
KW toluene monooxygenase; TMO; pcu gene; p-cresol.
XX
XX Pseudomonas mendocina KR-1.
XX
XX WO200192539-A2.
XX

```



```

PD 06-DEC-2001.
XX
PF 22-MAY-2001; 2001WO-US16574.
XX
XX 01-JUN-2000; 2000US-0585174.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX Ben-Bassat A, Cattermole M, Gatenby AA, Gibson KJ;
PI Ramos-Gonzales MI, Ramos-JL, Sarlasiani S;
XX
XX WPI; 2002-171436/22.
DR N-PSDB; ABA97241.
XX
XX New nucleic acid fragments encoding bacterial toluene monooxygenase
PT enzyme pathway, useful for isolating genes encoding proteins from the
PT same or other microbial species, and for producing para-hydroxybenzoate
PT
XX
XX Example 6; Page 90-91; 97pp; English.
XX
XX The invention relates to an isolated nucleic acid fragment encoding a
CC bacterial toluene monooxygenase enzyme pathway. The nucleic acid
CC fragments of the invention may be used to isolate genes encoding
CC proteins from the same or other microbial species. Bacterial strains
CC transformed with the p-cresol utilizing (pcu) genes are useful for
CC producing para-hydroxybenzoate (PHBA) which can be used for
CC synthesizing liquid crystal polymers (LCP). The current sequence
CC represents the Tmox amino acid sequence.
XX
XX Sequence 456 AA;
SQ
    Query Match      13.9%; Score 133.5; DB 23; Length 456;
    Best Local Similarity 26.7%; Pred. No. 7.6e-06;
    Matches 40; Conservative 26; Mismatches 75; Indels 9; Gaps 4;

QY 13 GSUTLK---LPAYWELSGFHQLTDQWAIHYSKYKTEWSR-FKELRGKYQDGSYEAFTKK 68
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 291 GDIRKDFEMPASITFCFAHQFNERLWLAADVRVYVMDVSDVDFKSGGDIIDIELP 350

QY 69 EEYKDNRSFAIGTYTSLNDALTRAGLAYDKAAKTHLSAS-IPDTRMYSIGATYKET 127
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 351 HNYQDITVASIGYAIRVNDKLRAGYSYAAQALDSRLILPVIPAYLKKHVSIGSDYSFD 410

QY 128 P-----NLSVDVGVFAHLRGKKKKHFEVTQNIK 153
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 411 KSKLNLAIISFGLKESLNTPSYLSGTETLK 440

RESULT 8
AAR40060
ID AAR40060 standard; peptide; 35 AA.
XX
AC AAR40060;
XX
XX 04-FEB-1994 (first entry)
DT
XX
XX Hib OMP p1 peptide Hibp1-8 (279-312).
DE
XX
XX Haemophilus influenzae; type b; Hib; outer membrane protein; P1; P2;
KW P6; vaccine; antibody; detection; lipoglycopeptide conjugate;
KW Immunogen.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 35
FT /note= "May be absent"
XX
XX WO9315205-A.
PN
XX
XX 05-AUG-1993.
PD

```

---

```

PF 03-FEB-1993; 93WO-CA00041.
XX
XX 03-FEB-1992; 92GB-0002219.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Chong P, Kandil A, Klein MII, Sia C;
PI WPI; 1993-258681/32.
XX
XX Synthetic Haemophilus influenzae conjugate vaccine - comprising
PT T-helper cell determinants and B-cell epitope(s) linked to
PT synthetic oligo:saccharide(s)
XX
XX Table 1; Page 47; 99pp; English.
XX
XX The sequences given in AAR40053-101 are peptide fragments derived from
CC the Haemophilus influenzae type b (Hib) outer membrane proteins P1,
CC P2 and P6. These peptides may be used in a vaccine against Hib
CC infection and antibodies against these peptides may be used in test
CC kits to detect H. influenzae in a sample. The vaccine may further
CC comprise an immunogenic or immunostimulatory molecule or the peptides
CC may be modified with lipids, or linked to synthetic PRP as synthetic
CC lipoglycopeptide conjugates to produce alternative vaccines.
XX
XX Sequence 35 AA;
SQ
    Query Match      13.5%; Score 129; DB 14; Length 35;
    Best Local Similarity 66.7%; Pred. No. 6.6e-07;
    Matches 22; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 24 ELSGFHQLTDQWAIHYSKYKTEWSR-FKELRGKY 56
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 2 ELSGFHQLTDKLAHVYSKYKTHWSRLTKLHASF 34

RESULT 9
AAR12452
ID AAR12452 standard; Protein; 33 AA.
XX
AC AAR12452;
XX
XX 26-JUL-1991 (first entry)
DT
XX
XX Hibp1-10(339-370) H. influenzae Type b outer membrane protein epitope.
DE
XX
XX bacterial meningitis; vaccine; P1 gene; T-cell antigen; OMP.
KW
XX
XX Haemophilus influenzae.
OS
XX
XX WO9106652-A.
PN
XX
XX 16-MAY-1991.
PD
XX
XX 31-OCT-1990; 90WO-CA00374.
PF
XX
XX 31-OCT-1989; 89GB-0024473.
PR
XX
XX (CONN-) CONNAUGHT LAB LTD.
PA
XX (UNIW ) WASHINGTON UNIV ST LOUIS.
XX
XX Munson RS, Grass S, Chong P, Yang Y, Fahlm R, McVerry P;
PI Klein M;
XX
XX WPI; 1991-164201/22.
DR
XX
XX Outer membrane protein of Haemophilus influenzae type B - used as
PT vaccine against infections, esp. in infants and for diagnosis
PT
XX
XX Disclosure; Table 2; 33pp; English.
PS
XX
XX Synthetic peptide corresponding to amino acids 339 to 370 of the H.
CC influenzae P1 protein. The C-terminal Cys residue is additional and
CC

```

CC allows coupling to a carrier in one specific orientation. The  
 CC free peptide was used as an immunogen to produce antibodies showing  
 CC that the peptide comprises both T-helper determinant and B-cell  
 CC epitope(s).  
 CC See also AAR12446-R12451 and AAR12453-5 and AAR12083.

XX Sequence 33 AA;

Query Match 11.6%; Score 111; DB 12; Length 33;  
 Best Local Similarity 71.0%; Pred. No. 5,7e-05;  
 Matches 22; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 85 LNDALTLRAGLAYDKASKTHLSASIPDTR 115

Db 1 LYEKLTLRAGIAYDQAASRHRSAIPDTR 31

# RESULT 10

AAR40062  
 ID AAR40062 standard; peptide: 33 AA.

XX AC AAR40062;

XX DT 04-FEB-1994 (first entry)

XX DE Hib OMP P1 peptide HIBP1-10 (339-370).

XX KW Haemophilus influenzae; type b; Hib; outer membrane protein; P1; P2;  
 KW P6; vaccine; antibody; detection; lipoglycopeptide conjugate;  
 KW immunogen.

XX OS Synthetic.

XX FH Key Location/Qualifiers  
 FT Misc-difference 33  
 ET /note= "May be absent"

XX W09315205-A.

XX PN 05-AUG-1993.

XX PF 03-FEB-1993; 93WO-CA00041.

XX PR 03-FEB-1992; 92GB-0002219.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX PI Chong P, Kandil A, Klein MH, Sia C;

XX DR WPI; 1993-258681/32.

XX PT Synthetic Haemophilus influenzae conjugate vaccine - comprising  
 PT T-helper cell determinants and B-cell epitope(s) linked to  
 PT synthetic oligo:saccharide(s)

XX PS Table 1; Page 47; 99pp; English.

XX CC The sequences given in AAR40053-101 are peptide fragments derived from  
 CC the Haemophilus influenzae type b (Hib) outer membrane proteins: P1,  
 CC P2 and P6. These peptides may be used in a vaccine against Hib  
 CC infection and antibodies against these peptides may be used in test  
 CC kits to detect H. influenzae in a sample. The vaccine may further  
 CC comprise an immunogenic or immunostimulatory molecule or the peptides  
 CC may be modified with lipids, or linked to synthetic Pp as synthetic  
 CC lipoglycopeptide conjugates to produce alternative vaccines.

XX Sequence 33 AA;

Query Match 11.6%; Score 111; DB 14; Length 33;  
 Best Local Similarity 71.0%; Pred. No. 5,7e-05;  
 Matches 22; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 85 LNDALTLRAGLAYDKASKTHLSASIPDTR 115

Db 1 LYEKLTLRAGIAYDQAASRHRSAIPDTR 31

# RESULT 11

AAB46323  
 ID AAB46323 standard; Protein: 587 AA.

XX AC AAB46323;

XX DT 05-APR-2001 (first entry)

XX DE H. pylori HPS188 protein.

XX KW Microbial infection; antibacterial; Helicobacter pylori infection;  
 KW vaccine; screening.

XX OS Helicobacter pylori.

XX PN W0200073502-A2.

XX PD 07-DEC-2000.

XX PF 31-MAY-2000; 2000WO-EP05024.

XX PR 31-MAY-1999; 99DE-1024965.

XX PR 17-JUN-1999; 99DE-1027740.

XX PR 21-JUL-1999; 99DE-1034029.

XX PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX PA (CREA-) CREATOGEN GMBH.

XX PI Apfel H, Fuchs TM, Gibbs CP, Hueck CJ, Meyer TF;

XX DR WPI; 2001-049948/06.

XX DR N-PSDB; AAF25600.

XX PT Preparing an agent for diagnosis or control of microbial infection,  
 PT useful particularly against Helicobacter, based on identification of  
 PT essential genes in defective mutants -

XX PS Claim 37; Page 274-276; 366pp; German.

XX CC This invention describes a novel preparation of an agent (A) for  
 CC detection, prevention and/or treatment of microbial infection by:  
 CC (i) identifying essential genes (I) and corresponding polypeptides  
 CC (II); (ii) identifying compounds that are directed against (II) and  
 CC inactivate the microbe; (iii) testing these for suitability for use; and  
 CC (iv) formulating selected (A). Identifying essential genes (I) comprises  
 CC preparation of gene-deficient microorganisms by conditional antisense  
 CC inhibition (CAI) and/or subtractive recombination mutagenesis (SRM),  
 CC then determining viability and/or survival of the deficient organisms.  
 CC The products of the invention have antibacterial activity. (A) (which may  
 CC be a nucleic acid (Ia), vector or host cell containing (Ia), derived  
 CC polypeptide (Iia), or fragments, (Iia)-specific antibodies or their  
 CC fragments or an inhibitor of (Iia)) are particularly used for diagnosis,  
 CC treatment or prevention of infection by Helicobacter pylori. Particularly  
 CC (Ia) and (Iia) are used in DNA, subunit or live vaccines. The method  
 CC identifies essential genes, including those that have homologs in other  
 CC species, so identified (A) should have a broad spectrum of activity. Many  
 CC gene-deficient cells can be screened quickly, in an automated process,  
 CC and the identified genes can be used for screening without purification.

XX SQ Sequence 587 AA;

Query Match 11.3%; Score 108; DB 22; Length 587;  
 Best Local Similarity 23.3%; Pred. No. 0.0067;  
 Matches 49; Conservative 29; Mismatches 78; Indels 54; Gaps 10;

OY 6 LCPYTG----KGSJTL--KLPAWELSGFHO-LTDQWAIHYSKYKTEWSRKFEL----- 52

Db 396 LCPSLGSLVLTGKSLNINVSLOTLSLAYAHQFFKDLRIEGVFERTFWSQGNKFLVTDF 455



XX





```

XX PR 11-APR-2000; 2000FR-0004629.
XX PA (INSP ) INST PASTEUR.
XX PI Buchrieser C, Frangeul L, Couve E, Rusnlok C, Fsihi H, Dehoux P;
XX PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
XX PI Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;
XX PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
XX PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
XX PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
XX PI Maduenio E, De Pablos B, Wehland J, Kaeist U, Entlan K, Hauf J;
XX PI Rose M, Voss H;
XX DR WPI; 2002-010914/01.
XX PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX PT and prevention of Listeria and related bacterial infections, and
XX PT related polypeptides
XX PS Claim 6; SEQ ID NO 32; 192pp; French.
XX CC The present invention relates to the genome sequence of Listeria
XX CC monocytogenes Ecbe (see ABA03041). The genome sequence and fragments of
XX CC it are useful for selecting probes and primers for detecting genes in L.
XX CC monocytogenes and related organisms, and for studying genetic
XX CC polymorphisms and other genomes. The present sequence is a protein
XX CC encoded by the genome sequence of the present invention. Proteins
XX CC expressed from the genome sequence are useful for raising specific
XX CC antibodies, identification of L. monocytogenes and related organisms, and
XX CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
XX CC B12. The genome sequence and proteins encoded by it are also useful for
XX CC selecting compounds that regulate gene expression and cell replication
XX CC and modulate L. monocytogenes-related diseases. In addition, the genome
XX CC sequence and proteins encoded by it are useful in pharmaceutical and
XX CC vaccines compositions for the treatment or prevention of infections by L.
XX CC monocytogenes and related organisms.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 572 AA;

Query Match 9.8%; Score 94; DB 23; Length 572;
Best Local Similarity 26.1%; Pred. No. 0.22;
Matches 57; Conservative 30; Mismatches 67; Indels 64; Gaps 12;

QY 4 GVLGPIYIGKSLTLKLPAYWELSGFHQLTDQWAIHYKYKTEWSRPFKELR-GKYQDCSGY 62
DB 140 GIKGSYNGK-SVTMKTWEYSDSGWYQINAFKYPKSHKESLEDNAKRLRNGPSWDSYY 198
QY 63 EAFTKK--EEYKD-----NSRFAIGTTSY--LN-----D 87
DB 199 KGARENAKYKDATANLQRYATDNTYASKLNTLTQYDTLYDTIKQKNVSED 258
QY 88 ALTURA-----GLAYDKAASKTHLSASIPDTR-----MMWYSIGATYKFTPNLSVD 133
DB 259 AKVVKADGCHGVYSGIYNTSAASAKKLTSGAPYNNKDKILKEGTTSRGTWVQFSLNNKV- 317
QY 134 VGFALRGKKKHVF---ETQNIKGLLLVEADYTTTKATA 168
DB 318 IGWM-----DKRAFYYPKATNKVTKLNL-----TGKITA 346

RESULT 21
AAB07698
ID AAB07698 standard; Protein; 691 AA.
XX AC AAB07698;
XX XX
XX XX 07-NOV-2000 (first entry)
XX DE A Neisseria meningitidis BASB053 polypeptide.

XX XX BASB053; Neisseria meningitidis infection; vaccine.
XX OS Neisseria meningitidis.
XX PN W0200042193-A1.
XX PD 20-JUL-2000.
XX PF 10-JAN-2000; 2000WO-EP00137.
XX PR 15-JAN-1999; 99GB-0000959.
XX PR 28-JAN-1999; 99GB-0001903.
XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX PI Ruelle J;
XX DR WPI; 2000-476062/41.
XX DR N-PSDB; AAA59217.
XX PT New Neisseria meningitidis polypeptide useful for diagnosis of
XX PT Neisseria infection and for development of vaccines against such
XX PT infection
XX PS Claim 3; Page 56; 92pp; English.
XX CC The present sequence represents a Neisseria meningitidis BASB053
XX CC polypeptide. The BASB053 polypeptide, or an antibody immunospecific
XX CC for BASB053 may be identified in a biological sample in order to
XX CC diagnose a Neisseria meningitidis infection in an animal. The BASB053
XX CC polypeptides and polynucleotides may be used as vaccines, for
XX CC generating an immune response in an animal. A composition comprising
XX CC at least one antibody immunospecific for BASB053 may be used to
XX CC treat humans infected with Neisseria meningitidis.
XX SQ Sequence 691 AA;

Query Match 9.8%; Score 93.5; DB 21; Length 691;
Best Local Similarity 23.5%; Pred. No. 0.32;
Matches 43; Conservative 23; Mismatches 70; Indels 47; Gaps 9;

QY 6 LGPIYIGKSLTLKLPAYWELSGFHQLTDQWAIHYKYKTEWSRPFKELRGKYQDCSGY 65
DB 547 LGKRVMEGVET-----EISG--AVTPKQIHAGYSYLH-SQIKTASNSRDDG---IF 592
QY 66 TKREYKDNSRFAIGTTSYSLNDALTLRAGLAYDKAASKTHLSASIPDTRMMWYSIGATYK 125
DB 593 LLMPKHSAN-----LWTTTYQVTPTELTIGGV---NAMSGITSSAGMHAGGYATFDMAAYR 645
QY 126 FTPNLSVDVGFALRGKKKHVEVETQNIKGLLLVEADYTTKATANLYGL-----NLN 176
DB 646 FTPKLLQINADNI--FNRHY-----YARVGGANTFIPGERSLTLNLR 688
QY 177 YRF 179
DB 689 YSF 691

RESULT 22
AAM50721
ID AAM50721 standard; Protein; 344 AA.
XX AC AAM50721;
XX XX
XX XX 18-APR-2002 (first entry)
XX DE Haemophilus paragallinarum strain 0222 haemagglutinin.
XX KW Haemagglutinin; HagA; antigen; vaccine; immunisation; coryza;
XX OS Haemophilus paragallinarum.

```

```
XX FH Key Location/Qualifiers
XX FT Peptide 1..21
XX FT /label= Signal_peptide
XX FT Protein 22..344
XX FT /label= Mature_protein
XX PN WO200204485-A1.
XX PD 17-JAN-2002.
XX PF 06-JUL-2001; 2001WO-AU00822.
XX PR 07-JUL-2000; 2000AU-0008652.
XX PA (UYQU ) UNIV QUEENSLAND.
XX PI Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;
XX DR WPI; 2002-154917/20.
XX DR N-PSDB; ABA91425.
XX PT New haemagglutinin polypeptide isolated from Haemophilus paragallinarum
XX PT useful as a vaccine for immunising chickens against coryza caused by
XX PT the species
XX PS Claim 2; Fig 4; 67pp; English.
XX CC The present sequence is that of the HagA haemagglutinin of
XX CC Haemophilus paragallinarum strain 0222 (serovar B), the causative
XX CC agent of infectious coryza in chickens. The invention provides
XX CC recombinant haemagglutinin polypeptides (see AAM50716-27) and
XX CC encoding nucleic acids (see ABA91417-27) from 11 strains (serovars
XX CC A, B and C) of H. paragallinarum, the causative agent of infectious
XX CC coryza of chickens. The polypeptides are useful in vaccines for
XX CC immunisation against infectious coryza, as are the nucleic acids
XX CC when expressed in attenuated bacteria, especially Salmonella or
XX CC Mycoplasma (claimed). The recombinant polypeptide is preferably
XX CC the mature protein, or a biologically active fragment, variant or
XX CC derivative, that is capable of eliciting an immune response,
XX CC providing protection against one or more strains of H.
XX CC paragallinarum in chickens. Also claimed are methods of using the
XX CC haemagglutinin polypeptides and nucleic acids for detection and
XX CC diagnosis of infectious coryza in chickens.
XX SQ Sequence 344 AA;
Query Match 9.7%; Score 93; DB 23; Length 344;
Best Local Similarity 20.7%; Pred. No. 0.14;
Matches 38; Conservative 31; Mismatches 71; Indels 44; Gaps 7;
QY 16 TLKLPAYWELSGFHQLTDQWAIHYSY-----KYTEWSRFKELRGKYODG 59
DB 55 TLRNSVTYGVFGGYQITDNFAVELGYDDFGRAKRQGGTIVIKYTNHGAHLSLKASYPVL 114
QY 60 SCYEAFK-----KEEYKDNSRFAIGT--YSLNDALTLAGLAYDKAASKTHLSASIP 111
DB 115 EGLDVIARVGAALIRSDYKPTKRAAPNQTHEHSLKVSVPFAGGLEYN-----LPSLP 166
QY 112 D-----TDRMWS-----IGATYKFTPNL-SVDVGFAPHLRKKKHFVETONIKGLLVE 159
DB 167 ELALRVEYQWNVKVRVEKDGSRVDYTPSIGSVTAGLSYRFQGSAPVPEPKVAKTFALN 226
QY 160 ADYT 163
DB 227 SDVT 230
RESULT 23
AAM50724
ID AAM50724 standard; Protein; 344 AA.
XX AC AAM50724;
```

```
XX 18-APR-2002 (first entry)
XX Haemophilus paragallinarum strain H-18 haemagglutinin.
XX Haemagglutinin; HagA; antigen; vaccine; immunisation; coryza;
XX chicken.
XX OS Haemophilus paragallinarum.
XX FH Key Location/Qualifiers
XX FT Peptide 1..21
XX FT /label= Signal_peptide
XX FT Protein 22..344
XX FT /label= Mature_protein
XX PN WO200204485-A1.
XX PD 17-JAN-2002.
XX PF 06-JUL-2001; 2001WO-AU00822.
XX PR 07-JUL-2000; 2000AU-0008652.
XX PA (UYQU ) UNIV QUEENSLAND.
XX PI Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;
XX DR WPI; 2002-154917/20.
XX DR N-PSDB; ABA91425.
XX PT New haemagglutinin polypeptide isolated from Haemophilus paragallinarum
XX PT useful as a vaccine for immunising chickens against coryza caused by
XX PT the species
XX PS Claim 2; Fig 4; 67pp; English.
XX CC The present sequence is that of the HagA haemagglutinin of
XX CC Haemophilus paragallinarum strain H-18 (serovar C), the causative
XX CC agent of infectious coryza in chickens. The invention provides
XX CC recombinant haemagglutinin polypeptides (see AAM50716-27) and
XX CC encoding nucleic acids (see ABA91417-27) from 11 strains (serovars
XX CC A, B and C) of H. paragallinarum, the causative agent of infectious
XX CC coryza of chickens. The polypeptides are useful in vaccines for
XX CC immunisation against infectious coryza, as are the nucleic acids
XX CC when expressed in attenuated bacteria, especially Salmonella or
XX CC Mycoplasma (claimed). The recombinant polypeptide is preferably
XX CC the mature protein, or a biologically active fragment, variant or
XX CC derivative, that is capable of eliciting an immune response,
XX CC providing protection against one or more strains of H.
XX CC paragallinarum in chickens. Also claimed are methods of using the
XX CC haemagglutinin polypeptides and nucleic acids for detection and
XX CC diagnosis of infectious coryza in chickens.
XX SQ Sequence 344 AA;
Query Match 9.7%; Score 93; DB 23; Length 344;
Best Local Similarity 20.7%; Pred. No. 0.14;
Matches 38; Conservative 31; Mismatches 71; Indels 44; Gaps 7;
QY 16 TLKLPAYWELSGFHQLTDQWAIHYSY-----KYTEWSRFKELRGKYODG 59
DB 55 TLRNSVTYGVFGGYQITDNFAVELGYDDFGRAKRQGGTIVIKYTNHGAHLSLKASYPVL 114
QY 60 SCYEAFK-----KEEYKDNSRFAIGT--YSLNDALTLAGLAYDKAASKTHLSASIP 111
DB 115 EGLDVIARVGAALIRSDYKPTKRAAPNQTHEHSLKVSVPFAGGLEYN-----LPSLP 166
QY 112 D-----TDRMWS-----IGATYKFTPNL-SVDVGFAPHLRKKKHFVETONIKGLLVE 159
DB 167 ELALRVEYQWNVKVRVEKDGSRVDYTPSIGSVTAGLSYRFQGSAPVPEPKVAKTFALN 226
QY 160 ADYT 163
```



```

Db      227 SDVT 230
      !!
RESULT 24
AAY75565
ID      AAY75565 standard; Protein; 154 AA.
AC      AAY75565;
XX
XX      21-MAR-2000 (first entry)
DT
XX
DE      Neisseria gonorrhoeae ORF 760 protein sequence SEQ ID NO:2604.
XX
KW      Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW      antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW      antibacterial; gene therapy.
XX
OS      Neisseria gonorrhoeae.
PN      WO9957280-A2.
XX
PD      11-NOV-1999.
XX
PF      30-APR-1999; 99WO-US09346.
XX
PR      01-MAY-1998; 98US-0083758.
PR      31-JUL-1998; 98US-0094869.
PR      02-SEP-1998; 98US-0098994.
PR      02-SEP-1998; 98US-0099062.
PR      09-OCT-1998; 98US-0103749.
PR      09-OCT-1998; 98US-0103794.
PR      09-OCT-1998; 98US-0103796.
PR      25-FEB-1999; 99US-0121528.
XX
PA      (CHIR ) CHIRON CORP.
PA      (GENO-) INST GENOMIC RES.
XX
XX      Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI      Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI      Tettelin H, Venter JC;
XX
XX      WPI: 2000-062150/05.
DR      N-PSDB; AA254327.
XX
PT      Novel Neisserial polypeptides predicted to be useful antigens for
PT      vaccines and diagnostics
PS      Claim 2; Page 1234; 1453pp; English.
XX
CC      AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AAY75941
CC      represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
CC      and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
CC      PCR primers used in the exemplification of the present invention. The
CC      polypeptides, the polynucleotides, antibodies and compositions of
CC      the invention can be used as vaccines, as diagnostic reagents, and as
CC      immunogenic compositions. The polypeptides can be used in the
CC      manufacture of medicaments for treating or preventing infection due to
CC      Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC      presence of Neisseria bacteria, or to raise antibodies. They may also
CC      be used to screen for agonists or antagonists, which may themselves
CC      have use as antibacterial agents. The polynucleotides of the invention
CC      may also be used in gene therapy protocols.
XX
SQ      Sequence 154 AA;

Query Match          9.7%; Score 92.5; DB 21; Length 154;
Best Local Similarity 25.6%; Pred. No. 0.051;
Matches 46; Conservative 21; Mismatches 72; Indels 41; Gaps 10;

QY      6 LGPYIGKSLTLKLPAYWELSGFHLTDOWAIHYSKYKTEWSRFLRGKYQDGSYEAF 65
      !! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! :
DB      10 LGRVMEGVET-----EISG--AITPKWQHAGYSYLH-SQIKTAANPRDDG----IF 55

```

---

```

QY      66 TKKEEYKDSRFAIGTYSINDALTLRAGLAYDKAASKTHLSASIPDTRMWYSIGATYK 125
      :: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      56 LLVPKHSAN-----LWTTYQVTPGLTVGGV---NANSGITSSAGMIAGGVATFDMAAYR 108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      126 FTPNLSVDYGVFAHLRGKKHF-----VETQNIKGLLLVLEADYTTKATANLYGLNLNYRF 179
      !!! : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      109 FTPKLLQINADNI--FNRHYARVGVGTNTFNIPG---SERSLTA-----NLRYSF 154
      !! : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 25
AAM50716
ID      AAM50716 standard; Protein; 344 AA.
XX
XX      AAM50716;
XX
XX      18-APR-2002 (first entry)
DT
XX
DE      Haemophilus paragallinarum strain 0083 haemagglutinin.
XX
KW      Haemagglutinin; HagA; antigen; vaccine; immunisation; coryza;
KW      chicken.
XX
OS      Haemophilus paragallinarum.
XX
FH      Key Location/Qualifiers
FT      Peptide 1..21
FT      Protein /label= Signal_peptide
FT      Protein 22..344 /label= Mature_protein
XX
XX      WO200204485-A1.
XX
XX      17-JAN-2002.
XX
XX      06-JUL-2001; 2001WO-AU00822.
XX
XX      07-JUL-2000; 2000AU-0008652.
XX      (UYQU ) UNIV QUEENSLAND.
XX
XX      Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;
XX      WPI: 2002-154917/20.
DR      N-PSDB; ABA91417.
XX
PT      New haemagglutinin polypeptide isolated from Haemophilus paragallinarum
PT      useful as a vaccine for immunising chickens against coryza caused by
PT      the species
XX
PS      Claim 2; Fig 4; 67pp; English.
XX
CC      The present sequence is that of the HagA haemagglutinin of
CC      Haemophilus paragallinarum strain 0083 (serovar A), the causative
CC      agent of infectious coryza in chickens. The invention provides
CC      recombinant haemagglutinin polypeptides (see AAM50716-27) and
CC      encoding nucleic acids (see ABA91417-27) from 11 strains (serovars
CC      A, B and C) of H. paragallinarum, the causative agent of infectious
CC      coryza of chickens. The polypeptides are useful in vaccines for
CC      immunisation against infectious coryza, as are the nucleic acids
CC      when expressed in attenuated bacteria, especially Salmonella or
CC      Mycoplasma (claimed). The recombinant polypeptide is preferably
CC      the mature protein, or a biologically active fragment, variant or
CC      derivative, that is capable of eliciting an immune response,
CC      providing protection against one or more strains of H.
CC      paragallinarum in chickens. Also claimed are methods of using the
CC      haemagglutinin polypeptides and nucleic acids for detection and
CC      diagnosis of infectious coryza in chickens.
XX
SQ      Sequence 344 AA;

Query Match          9.4%; Score 90; DB 23; Length 344;
Best Local Similarity 20.1%; Pred. No. 0.29;

```





```
PR 07-JUL-2000; 2000AU-0008652.
PA (UYOU ) UNIV QUEENSLAND.
PI Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;
XX WPI; 2002-154917/20.
DR N-PSDB; ABA91424.
XX New haemagglutinin polypeptide isolated from Haemophilus paragallinarum
PT useful as a vaccine for immunising chickens against coryza caused by
PT the species
XX Claim 2; Fig 4; 67pp; English.
XX The present sequence is that of the Hga haemagglutinin of
CC Haemophilus paragallinarum strain Modesto (serovar C), the causative
CC agent of infectious coryza in chickens. The invention provides
CC recombinant haemagglutinin polypeptides (see AAM50716-27) and
CC encoding nucleic acids (see ABA91417-27) from 11 strains (serovars
CC A, B and C) of H. paragallinarum, the causative agent of infectious
CC coryza of chickens. The polypeptides are useful in vaccines for
CC immunisation against infectious coryza, as are the nucleic acids
CC when expressed in attenuated bacteria, especially Salmonella or
CC Mycoplasma (claimed). The recombinant polypeptide is preferably
CC the mature protein, or a biologically active fragment, variant or
CC derivative, that is capable of eliciting an immune response,
CC providing protection against one or more strains of H.
CC paragallinarum in chickens. Also claimed are methods of using the
CC haemagglutinin polypeptides and nucleic acids for detection and
CC diagnosis of infectious coryza in chickens.
XX SQ Sequence 341 AA;
Query Match 9.4%; Score 89.5; DB 23; Length 341;
Best Local Similarity 20.4%; Pred. No. 0.33;
Matches 37; Conservative 32; Mismatches 71; Indels 41; Gaps 7;
QY 16 TLKLPAYWELSGFHOLTQOWATHYSY-----KYTEWSRFEKLRKYQDG 59
DB 55 TLRSVTYGVFGGYOITDNFAVELGYDDFGRAKFRQDGETVIKHTNHGAHLSKASYPVL 114
QY 60 SCYEAFK-----KEEKDNSRFAIGTT--YSLNDALTLAGLAYDKAASKTHLSASIP 111
DB 115 EGLDVYARVGAALIRSDYKPTKRAAPNETHEHSLKVPVFAAGGLEYN-----LPSLP 166
QY 112 D-----TDRMWS-----IGATYKFTPNL-SVDVGFALRGKKHFEVTONIKGLLVEADY 162
DB 167 ELALRVEYQWNVKVGROGSRVDYTPSIGSVTAGLSYRFQGSAPVVEPKVAKTFALNSDV 226
QY 163 T 163
DB 227 T 227
RESULT 30
AAM50726
ID AAM50726 standard; Protein; 341 AA.
XX AC AAM50726;
XX 18-APR-2002 (first entry)
XX Haemophilus paragallinarum strain HP60 haemagglutinin.
XX Haemagglutinin; Hga; antigen; vaccine; immunisation; coryza;
KW chicken.
XX Haemophilus paragallinarum.
XX Key Location/Qualifiers
FH 1..21
FT Peptide /label= Signal_peptide
FT
```

```
FT Protein 22..341
XX /label= Mature_protein
PN WO200204485-A1.
XX 17-JAN-2002.
XX 06-JUL-2001; 2001WO-AU00822.
XX 07-JUL-2000; 2000AU-0008652.
XX (UYOU ) UNIV QUEENSLAND.
PI Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;
XX WPI; 2002-154917/20.
DR N-PSDB; ABA91427.
XX New haemagglutinin polypeptide isolated from Haemophilus paragallinarum
PT useful as a vaccine for immunising chickens against coryza caused by
PT the species
XX Claim 2; Fig 4; 67pp; English.
XX The present sequence is that of the Hga haemagglutinin of
CC Haemophilus paragallinarum strain HP60 (serovar C), the causative
CC agent of infectious coryza in chickens. The invention provides
CC recombinant haemagglutinin polypeptides (see AAM50716-27) and
CC encoding nucleic acids (see ABA91417-27) from 11 strains (serovars
CC A, B and C) of H. paragallinarum, the causative agent of infectious
CC coryza of chickens. The polypeptides are useful in vaccines for
CC immunisation against infectious coryza, as are the nucleic acids
CC when expressed in attenuated bacteria, especially Salmonella or
CC Mycoplasma (claimed). The recombinant polypeptide is preferably
CC the mature protein, or a biologically active fragment, variant or
CC derivative, that is capable of eliciting an immune response,
CC providing protection against one or more strains of H.
CC paragallinarum in chickens. Also claimed are methods of using the
CC haemagglutinin polypeptides and nucleic acids for detection and
CC diagnosis of infectious coryza in chickens.
XX SQ Sequence 341 AA;
Query Match 9.4%; Score 89.5; DB 23; Length 341;
Best Local Similarity 20.4%; Pred. No. 0.33;
Matches 37; Conservative 32; Mismatches 71; Indels 41; Gaps 7;
QY 16 TLKLPAYWELSGFHOLTQOWATHYSY-----KYTEWSRFEKLRKYQDG 59
DB 55 TLRSVTYGVFGGYOITDNFAVELGYDDFGRAKFRQDGETVIKHTNHGAHLSKASYPVL 114
QY 60 SCYEAFK-----KEEKDNSRFAIGTT--YSLNDALTLAGLAYDKAASKTHLSASIP 111
DB 115 EGLDVYARVGAALIRSDYKPTKRAAPNETHEHSLKVPVFAAGGLEYN-----LPSLP 166
QY 112 D-----TDRMWS-----IGATYKFTPNL-SVDVGFALRGKKHFEVTONIKGLLVEADY 162
DB 167 ELALRVEYQWNVKVGROGSRVDYTPSIGSVTAGLSYRFQGSAPVVEPKVAKTFALNSDV 226
QY 163 T 163
DB 227 T 227
RESULT 31
AAR77432
ID AAR77432 standard; Protein; 459 AA.
XX AC AAR77432;
XX 22-JUN-1996 (first entry)
XX Moraxella catarrhalis outer membrane protein E.
DE
```

```
XX KW Bacterium; Branhamella catarrhalis; sinusitis; conjunctivitis;
KW pneumonia; endocarditis; septicemia; meningitis; otitis media;
KW lower respiratory tract infection; chronic bronchitis;
KW chronic obstructive pulmonary disease; vaccine; diagnostic;
KW immunoassay.
XX OS Moraxella catarrhalis.
XX FH Key
XX FT Peptide
XX FT 1..25
XX FT /label= sig_peptide
XX FT Region
XX FT 23..34
XX FT /note= "antigenic determinant hydrophilic region"
XX FT Region
XX FT 234..299
XX FT /note= "antigenic determinant hydrophilic region"
XX FT Region
XX FT 369..374
XX FT /note= "antigenic determinant hydrophilic region"
XX PN W09531215-AL.
XX PD 23-NOV-1995.
XX PE 20-APR-1995; 95W0-US05134.
XX PR 17-MAY-1994; 94US-0245758.
XX PA (UUNY ) UNIV NEW YORK STATE RES FOUND.
XX PI Bhushan R, Murphy TF;
XX PR 1996-010692/01.
XX DR N-PSDB; AAT04376.
XX RX Vaccine contg outer membrane protein E of Moraxella catarrhalis -
PT for the detection of M.catarrhalis-specific antisera and in
PT diagnostic immunoassays
XX PS Claim 11; Page 40-43; 58pp; English.
XX CC This E protein from M. catarrhalis strain 25240 may be produced
CC recombinantly from cells cultured from a host cell system engineered
CC to include a vector containing a nucleotide sequence that regulates
CC expression of DNA encoding E epitopes. The host cell system may be
CC bacteria, yeast, filamentous fungi, insect and mammalian cell lines.
CC The protein may be used to detect M. catarrhalis-specific antisera
CC in body fluid and may be used in a vaccine against M. catarrhalis.
CC It may also be used as an antigen in diagnostic immunoassays.
XX SQ Sequence 459 AA;
Query Match 9.3%; Score 89; DB 17; Length 459;
Best Local Similarity 23.2%; Pred. No. 0.57;
Matches 42; Conservative 19; Mismatches 84; Indels 36; Gaps 6;
QY 9 YICKGSLTKLPAYWELSGFHQLTDQWAIHYSKYKTEWSRPFKELRGKYQDGSYEAFKK 68
DB 292 YAGKDKFKVTLPSWNLDFQTVGNPTLLTAKVRYVPWSDFDIRPQYTE-----TTK 344
QY 69 EEEY-----KNSRFAIGTYSINDALTLRAGLAYDKAASKTHLSASIPDTRMWY 118
DB 345 LRYPOGLPIISYDKDQWSAEVGLKRVSDRLAVSGVGDGSGAGN---PASSLGPIKGY 401
QY 119 S---IGATYKFTPNLSVDVGFALHKKHFFVETQNTKGLLLVBEADYTTKATANLYGLNLN 176
DB 402 SLGLGARYNVTPEWSLSG-----GKYFKGDAQ-----AQLPTKDKVGNFDSNDG 447
QY 177 Y 177
DB 448 Y 448
RESULT 32
```

```
AYY31154
ID AAY31154 standard; Protein; 459 AA.
AC AAY31154;
XX DT 26-OCT-1999 (first entry)
XX DE M. catarrhalis outer membrane E protein.
XX KW Outer membrane protein; E protein; vaccine; antigen; epitope; therapy;
KW heat-modifiable protein; immunogen; prophylactic; active immunization;
KW antisera; passive immunization; prevention; otitis media; sinusitis;
KW conjunctivitis; lower respiratory tract infection; diagnosis;
XX KW immunoassay.
XX OS Moraxella catarrhalis.
XX FH Key
XX FT Peptide
XX FT 1..25
XX FT /label= signal_peptide
XX FT Protein
XX FT 26..459
XX PN US5948412-A.
XX PD 07-SEP-1999.
XX PE 03-MAR-1997; 97US-0810655.
XX PR 03-MAR-1997; 97US-0810655.
XX PR 17-MAY-1994; 94US-0245758.
XX PA (UUNY ) UNIV NEW YORK STATE RES FOUND.
XX PI Murphy TF;
XX DR WPI; 1999-517930/43.
XX DR N-PSDB; AA209296.
XX RX Antigenic peptide, oligopeptide and protein useful as vaccine for
PT Moraxella catarrhalis
XX PS Claim 1; Column 25-28; 20pp; English.
XX CC This invention describes a novel pure antigenic peptide, oligopeptide or
CC protein (I) with one or more epitopes of E, an outer membrane protein of
CC Moraxella catarrhalis. The E protein of the invention has an apparent
CC molecular weight of 35000-50000 daltons by sodium dodecyl
CC sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and is a
CC heat-modifiable protein. (I) can be used as immunogens in prophylactic
CC and/or therapeutic vaccine formulations for active immunization and for
CC generating protein-specific and peptide-specific antisera useful for
CC passive immunization. Antigenic formulations comprising (I) can be used
CC to prevent otitis media, sinusitis, conjunctivitis and lower respiratory
CC tract infections caused by Moraxella catarrhalis. (I) can be used as
CC antigens for diagnostic immunoassays. This sequence represents the
CC Moraxella catarrhalis strain 25240 outer membrane E protein which is
CC described in the method of the invention.
XX SQ Sequence 459 AA;
Query Match 9.3%; Score 89; DB 20; Length 459;
Best Local Similarity 23.2%; Pred. No. 0.57;
Matches 42; Conservative 19; Mismatches 84; Indels 36; Gaps 6;
QY 9 YICKGSLTKLPAYWELSGFHQLTDQWAIHYSKYKTEWSRPFKELRGKYQDGSYEAFKK 68
DB 292 YAGKDKFKVTLPSWNLDFQTVGNPTLLTAKVRYVPWSDFDIRPQYTE-----TTK 344
QY 69 EEEY-----KNSRFAIGTYSINDALTLRAGLAYDKAASKTHLSASIPDTRMWY 118
DB 345 LRYPOGLPIISYDKDQWSAEVGLKRVSDRLAVSGVGDGSGAGN---PASSLGPIKGY 401
QY 119 S---IGATYKFTPNLSVDVGFALHKKHFFVETQNTKGLLLVBEADYTTKATANLYGLNLN 176
```

Db 402 SIGLGARYNVTPWLSLG-----GKYKFGDAQ-----AQLPTKDKVGNFDSNDG 447  
 Qy 177 Y 177  
 Db 448 Y 448

RESULT 33  
 AAM50718  
 ID AAM50718 standard; Protein; 344 AA.  
 XX AAM50718;  
 DT 18-APR-2002 (first entry)  
 DE Haemophilus paragallinarum strain 2403 haemagglutinin.  
 KW Haemagglutinin; Hagg; antigen; vaccine; immunisation; coryza;  
 KW chicken.  
 XX Haemophilus paragallinarum.  
 OS  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /label= Signal\_peptide  
 FT Protein 22..344  
 FT /label= Mature\_protein  
 XX WO200204485-A1.  
 PN 17-JAN-2002.  
 PD  
 XX  
 XX 06-JUL-2001; 2001WO-AU00822.  
 PF  
 XX 07-JUL-2000; 2000AU-0008652.  
 PR  
 XX (UYQU ) UNIV QUEENSLAND.  
 PA  
 XX Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;  
 PI WPI; 2002-154917/20.  
 DR N-PSDB; ABA91419.  
 DR  
 XX New haemagglutinin polypeptide isolated from Haemophilus paragallinarum  
 PT useful as a vaccine for immunising chickens against coryza caused by  
 PT the species  
 PS Claim 2; Fig 4; 67pp; English.  
 XX  
 CC The present sequence is that of the Hagg haemagglutinin of  
 CC Haemophilus paragallinarum strain 2403 (serovar A), the causative  
 CC agent of infectious coryza in chickens. The invention provides  
 CC recombinant haemagglutinin polypeptides (see AAM50716-27) and  
 CC encoding nucleic acids (see ABA91417-27) from 11 strains (serovars  
 CC A, B and C) of H. paragallinarum, the causative agent of infectious  
 CC coryza of chickens. The polypeptides are useful in vaccines for  
 CC immunisation against infectious coryza, as are the nucleic acids  
 CC when expressed in attenuated bacteria, especially Salmonella or  
 CC Mycoplasma (claimed). The recombinant polypeptide is preferably  
 CC the mature protein, or a biologically active fragment, variant or  
 CC derivative, that is capable of eliciting an immune response,  
 CC providing protection against one or more strains of H.  
 CC paragallinarum in chickens. Also claimed are methods of using the  
 CC haemagglutinin polypeptides and nucleic acids for detection and  
 CC diagnosis of infectious coryza in chickens.  
 XX SQ Sequence 344 AA;

Query Match 9.2%; Score 88; DB 23; Length 344;  
 Best Local Similarity 20.1%; Pred. No. 0.49;  
 Matches 37; Conservative 32; Mismatches 71; Indels 44; Gaps 7;

Qy 16 TLKLPAYWELSGFHLTDQWAIHYSY-----KYTEWSRFKELRGKYODG 59  
 Db 55 TLRSVTVGVGGYQITDNFAVELGVDGGRAKSRQGGTIVIKHTNHGAHLSLKASYPVL 114  
 Qy 60 SGYEAFK-----KEEYKDNSRFAIGTT--YSLNDALTLRAGLAYDKAASKTHLSASIP 111  
 Db 115 EGLDVYARVGAALIRSDYKPTKRAAPNOTHEHLSKVPFVAGGLEYN-----LPSLP 166  
 Qy 112 D-----TDRMWSY-----IGATYKFTPNL--SVDVGFHRLGKKKHVFETONIKGLLVE 159  
 Db 167 ELALRVEYQWYKVRVEKDGSRVDYTPSICSVTAGLSYRFGQSAPVVEPKVAKTFALN 226  
 Qy 160 ADYT 163  
 Db 227 SDVT 230

RESULT 34  
 AAM50722  
 ID AAM50722 standard; Protein; 344 AA.  
 XX AAM50722;  
 DT 18-APR-2002 (first entry)  
 DE Haemophilus paragallinarum strain 2671 haemagglutinin.  
 KW Haemagglutinin; Hagg; antigen; vaccine; immunisation; coryza;  
 KW chicken.  
 XX Haemophilus paragallinarum.  
 OS  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /label= Signal\_peptide  
 FT Protein 22..344  
 FT /label= Mature\_protein  
 XX WO200204485-A1.  
 PN 17-JAN-2002.  
 PD  
 XX  
 XX 06-JUL-2001; 2001WO-AU00822.  
 PF  
 XX 07-JUL-2000; 2000AU-0008652.  
 PR  
 XX (UYQU ) UNIV QUEENSLAND.  
 PA  
 XX Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;  
 PI WPI; 2002-154917/20.  
 DR N-PSDB; ABA91423.  
 DR  
 XX New haemagglutinin polypeptide isolated from Haemophilus paragallinarum  
 PT useful as a vaccine for immunising chickens against coryza caused by  
 PT the species  
 PS Claim 2; Fig 4; 67pp; English.  
 XX  
 CC The present sequence is that of the Hagg haemagglutinin of  
 CC Haemophilus paragallinarum strain 2671 (serovar B), the causative  
 CC agent of infectious coryza in chickens. The invention provides  
 CC recombinant haemagglutinin polypeptides (see AAM50716-27) and  
 CC encoding nucleic acids (see ABA91417-27) from 11 strains (serovars  
 CC A, B and C) of H. paragallinarum, the causative agent of infectious  
 CC coryza of chickens. The polypeptides are useful in vaccines for  
 CC immunisation against infectious coryza, as are the nucleic acids  
 CC when expressed in attenuated bacteria, especially Salmonella or  
 CC Mycoplasma (claimed). The recombinant polypeptide is preferably  
 CC the mature protein, or a biologically active fragment, variant or  
 CC derivative, that is capable of eliciting an immune response,  
 CC providing protection against one or more strains of H.  
 CC paragallinarum in chickens. Also claimed are methods of using the  
 CC haemagglutinin polypeptides and nucleic acids for detection and  
 CC diagnosis of infectious coryza in chickens.

CC haemagglutinin polypeptides and nucleic acids for detection and  
CC diagnosis of infectious coryza in chickens.

XX SQ Sequence 344 AA;

Query Match 9.2%; Score 88; DB 23; Length 344;  
Best Local Similarity 20.1%; Pred. No. 0.49;  
Matches 37; Conservative 32; Mismatches 71; Indels 44; Gaps 7;

QY 16 TLKLPAYWELSGPHLTQDWAHYSY-----KYTEWSRFEKELRGKQYQDG 59  
DB 55 TLRNVTYGVFGQITDNFAVELGVDYDFGRARRGGTIVIRKTHNGAHLSLKASTPYVL 114  
QY 60 SGYEAFK-----KEEYKNSRFAIGTT--YSLNDALTLRAGLAYDKAAASKTHLSASIP 111  
DB 115 EGLDVYARVGAALIRSDYKPTKRAAPNQTHEHSLKVSVPFAGGLEYN-----LPSLP 166  
QY 112 D-----TDRMYS-----IGATYKFTPNL-SVDVGFALHRLGKKHFEVTONIKGLLVE 159  
DB 167 ELALRVEYQWNVKVRVEKDGSRVDYTPSIGSVTAGLSYRFGOSAPVVEPKVAKTFALN 226  
QY 160 ADVT 163  
DB 227 SDVT 230

RESULT 35

AA12455  
ID AAR12455 standard; Protein; 36 AA.

XX AC AAR12455;

XX DT 26-JUL-1991 (first entry)

XX DE HIBP1-14(400-437) 6U H. influenzae Type b OMP epitope.

XX KW bacterial meningitis; vaccine; Pl gene; T-cell antigen;  
KW outer membrane protein.

XX OS Haemophilus influenzae.

XX PN WO9106652-A.

XX PD 16-MAY-1991.

XX PF 31-OCT-1990; 90WO-CA00374.

XX PR 31-OCT-1989; 89GB-0024473.

XX PA (CONN-) CONNAUGHT LAB LTD.  
(UNIW ) WASHINGTON UNIV ST LOUIS.

XX PI Munson RS, Grass S, Chong P, Yang Y, Fahim R, McVerry P;  
PI Klein M;

XX DR WPI; 1991-164201/22.

XX PT Outer membrane protein of Haemophilus influenzae type B - used as  
PT vaccine against infections, esp. in infants and for diagnosis

XX PS Disclosure; Table 2; 33pp; English.

XX CC Synthetic peptide corresponding to amino acids 400 to 437 of the H.  
CC influenzae strain 8358 (OMP subtype 6U) pl protein. The N-terminal  
CC Cys residue is additional and allows coupling to a carrier in one  
CC specific orientation. The free peptide was used as an immunogen to  
CC produce antibodies showing that the peptide comprises both T-helper  
CC determinant and B-cell epitope(s).  
CC See also AAR12446-R12454 and AAQ12083.

XX SQ Sequence 36 AA;

Query Match 9.1%; Score 87.5; DB 12; Length 36;

Best Local Similarity 54.3%; Pred. No. 0.024;  
Matches 19; Conservative 4; Mismatches 11; Indels 1; Gaps 1;  
QY 146 FVETQNIK-GLLLEADYTTKATANLYGLNLNRYF 179  
DB 2 FKEAQQAAAGGFITTTANTYTSQAHANLYGLNLNYSF 36

RESULT 36

AA40066  
ID AAR40066 standard; peptide; 36 AA.

XX AC AAR40066;

XX DT 04-FEB-1994 (first entry)

XX DE HIB.OMP Pl peptide HIBP1-14 (400-437).

XX KW Haemophilus influenzae; type b; Hib; outer membrane protein; Pl; P2;  
KW P6; vaccine; antibody; detection; lipoglycopeptide conjugate;  
KW immunogen.

XX OS Synthetic.

XX FH Key Location/Qualifiers  
XX FT Misc-difference 1  
XX FT /note- "May be absent"

XX FT Misc-difference 5...14  
XX FT /note- "Residues are different to those found in Pl  
protein of H. influenzae strain 1H"

XX PN WO9315205-A.

XX PD 05-AUG-1993.

XX PF 03-FEB-1993; 93WO-CA00041.

XX PR 03-FEB-1992; 92GB-0002219.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX PI Chong P, Kandil A, Klein MH, Sia C;

XX DR WPI; 1993-258681/32.

XX PT Synthetic Haemophilus influenzae conjugate vaccine - comprising  
PT T-helper cell determinants and B-cell epitope(s) linked to  
PT synthetic oligo:saccharide(s)

XX PS Table 1; Page 48; 99pp; English.

XX CC The sequences given in AAR40053-101 are peptide fragments derived from  
CC the Haemophilus influenzae type b (Hib) outer membrane proteins Pl,  
CC P2 and P6. These peptides may be used in a vaccine against Hib  
CC infection and antibodies against these peptides may be used in test  
CC kits to detect H. influenzae in a sample. The vaccine may further  
CC comprise an immunogenic or immunostimulatory molecule or the peptides  
CC may be modified with lipids, or linked to synthetic PRP as synthetic  
CC lipoglycopeptide conjugates to produce alternative vaccines.

XX SQ Sequence 36 AA;

Query Match 9.1%; Score 87.5; DB 14; Length 36;  
Best Local Similarity 54.3%; Pred. No. 0.024;  
Matches 19; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

QY 146 FVETQNIK-GLLLEADYTTKATANLYGLNLNRYF 179  
DB 2 FKEAQQAAAGGFITTTANTYTSQAHANLYGLNLNYSF 36

RESULT 37

AAM50725





PS Claim 2; Fig 4; 67pp; English.

XX The present sequence is a consensus of the HagA haemagglutinins  
 CC of 11 strains of Haemophilus paragallinarum serovars A, B and C  
 CC (see AAM50716-26). The invention provides recombinant haemagglutinin  
 CC polypeptides and encoding nucleic acids (see ABA91417-27) from H.  
 CC paragallinarum, the causative agent of infectious coryza of chickens.  
 CC The polypeptides are useful in vaccines for immunisation against  
 CC infectious coryza, as are the nucleic acids when expressed in  
 CC attenuated bacteria, especially Salmonella or Mycoplasma (claimed).  
 CC The recombinant polypeptide is preferably the mature protein, or a  
 CC biologically active fragment, variant or derivative, that is capable  
 CC of eliciting an immune response, providing protection against one or  
 CC more strains of H. paragallinarum in chickens. Also claimed are  
 CC methods of using the haemagglutinin polypeptides and nucleic acids  
 CC for detection and diagnosis of infectious coryza in chickens.

XX Sequence 344 AA;

Query Match 9.0%; Score 86; DB 23; Length 344;

Best Local Similarity 20.1%; Pred. No. 0.81;

Matches 37; Conservative 31; Mismatches 72; Indels 44; Gaps 7;

OY 16 TLKLPAYWELSGFHLTDQWAIHVS-----KYTEWSRKEKLGKXQDG 59  
 Db 55 TLRSVTYGVGGYQITDNFAVELGVDYDFGRAXKXRGXETVXXXTNHGAHLSKASYPVL 114  
 OY 60 SGYEATK-----KEEYKDNSRFAIGTT--YSLNDALTLRAGLAYDKAASKTHLSASIP 111  
 Db 115 EGLDVARVGAALIRSDYKPTKRAAPNXTHEHSKVSVPVAGGLEYN-----LPSLP 166  
 OY 112 D-----TDRMYS-----ICATYKTPNL-SVDVGFALHURGKKKHVFETONIKGULLVE 159  
 Db 167 ELALRVEYQWNVKVRXXXDGRVDVTPSGVTAGLSYRFGOSAPVPEKVVAKTFALN 226  
 OY 160 ADVT 163  
 Db 227 SDVT 230

RESULT 39

AAM92998

ID AAM92998 standard; protein; 340 AA.

XX AAM92998;

DT 19-MAY-1999 (first entry)

DE E. coli OmpF porin protein.

KW OmpF; porin; amphiphilic alpha-structure; beta-structure; GLUT;  
 KW membrane protein; glucose transporter protein; function.

XX Escherichia coli.

PN WO9618957-A1.

PD 20-JUN-1996.

PF 13-DEC-1995; 95WO-0516126.

PR 14-DEC-1994; 94US-0355844.

XX (UYCO ) UNIV COLUMBIA NEW YORK.

XX Cheung M, Czegledy F, Fischbarg J, Iserovich P;

PI Li J;

XX WPI; 1996-300839/30.

XX Predicting tendency to form amphiphilic alpha and beta structures -  
 PT using a novel algorithm to calculate values for subsequent graphical  
 PT analysis to predict protein structure

XX

PS Disclosure; Fig 2; 105pp; English.

XX This invention describes a novel method for predicting the tendency of  
 CC a protein to form either: (a) an amphiphilic alpha-structure or (b) an  
 CC amphiphilic beta-structure. The methods are used particularly for  
 CC predicting the structure of membrane proteins such as glucose transporter  
 CC proteins (GLUTs). They can be used to discern the function of proteins.  
 CC They can also be used for the rational design or identification of  
 CC compounds which interact with the proteins or to engineer proteins having  
 CC particular structures. This sequence represents an Escherichia coli  
 CC OmpF porin which is used to illustrate the method of the invention.

XX Sequence 340 AA;

Query Match 8.9%; Score 85.5; DB 17; Length 340;

Best Local Similarity 25.2%; Pred. No. 0.9;

Matches 36; Conservative 22; Mismatches 60; Indels 25; Gaps 6;

OY 4 GVLPGYVIGKSLTLKLPAYWELSGFHLTDQW-----IHYSKYKTEWSRKEKLRG 54  
 Db 186 GIVGAY-GAADRNLQEA--QPLGNGKKAQWATGLKYDANNIYLAANYGETRNATPITN 242  
 OY 55 KYDGGSGYEAFYTKKEEYKDNSRFAIGTYSLNDALTLRAGLAYDKAASKTHLSASIPDPTD 114  
 Db 243 KFTNTSGFANKTQDVLVAQYQDFG-----LRPSIAYTKSKAKD--VEGIGDVD 290  
 OY 115 RM-WYSIGATYKFTPNLSVDVGF 136  
 Db 291 LVNVEFGATYYPNKNMSTIVDY 313

RESULT 40

AAB82612

ID AAB82612 standard; Protein; 341 AA.

XX AAB82612;

DT 02-OCT-2001 (first entry)

DE E. coli outer membrane protein ompF.

XX Outer membrane protein; ompF; structural protein; purification.

XX Escherichia coli.

PN WO200153333-A1.

PD 26-JUL-2001.

PF 01-NOV-2000; 2000WO-US30086.

PR 20-JAN-2000; 2000US-0490291.

XX (WELL/) MELLO C M.

PA (ARCI/) ARCIDIACONO S.

PA (BUTL/) BUTLER M M.

PA (USSA ) US SEC OF ARMY.

PI Mello CM, Arcidiacono S, Butler MM;

XX WPI; 2001-483136/52.

XX Recovering structural polypeptides in a biological sample, useful for  
 PT purifying and spinning spider silks and other structural proteins, an acid  
 PT comprises treating the sample containing the polypeptides with an acid

XX Claim 2; Page 43-44; 49pp; English.

XX The present sequence is that of the Escherichia coli ompF outer  
 CC membrane protein. The invention provides methods for purifying  
 CC structural proteins. Organic acids are used to lyse recombinant

CC cells or other biological samples (such as non-recombinantly  
CC derived cells), and enrich the purity and yields of structural  
CC proteins by hydrolysing many of the macromolecules while leaving  
CC the structural proteins intact. In the present case, lyophilised  
CC pellets of *E. coli* were lysed in 9.2 N valeric acid for 1 hr at  
CC room temperature. The cell lysate was clarified by centrifugation  
CC and applied to an SDS polyacrylamide gel for electrophoresis. The  
CC ompF protein was then blotted onto a nitrocellulose membrane for  
CC N-terminal sequencing. The simple extraction procedure yielded  
CC ompF in approximately 85% purity. The new method has the  
CC following advantages over prior art: it involves fewer steps,  
CC requires less time and smaller volumes of reagents, results in  
CC better recovery of protein at higher purity (70-99%), is easy to  
CC scale up, and fibres can be spun in an environmentally benign  
CC solution reducing hazardous waste accumulation and cost.  
XX  
SQ Sequence 341 AA:  
Query Match 8.9%; Score 85.5; DB 22; Length 341;  
Best Local Similarity 25.2%; Pred. NO. 0.9;  
Matches 36; Conservative 22; Mismatches 60; Indels 25; Gaps 6;  
QY 4 GVLGPEYICKGSLTKLPAYWELSGPHQLTDQWA-----IHYSKYTEWSREKELRG 54  
DB 186 GIVGAY-GAADRTNLQEA--OPLGNGKRAEQWATGLKYDANNIYLAANYGETRNATPITN 242  
QY 55 KYQDGSGYEAFTKKEEYKDNRSFAIGTGYSLNDALTLRAGLAYDKAASKTHLSASIPDTD 114  
DB 243 KFTNTSGFANKTQDVLVLAQYQDFG-----LRPSIAYTKSKAKD--VEGIGDVD 290  
QY 115 RM-WYSIGATYKFTPNLSVDVGF 136  
DB 291 LVNYPEVGATYYFNKNMSTYVDY 313

Search completed: May 12, 2003, 09:40:18  
Job time : 39 secs

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM protein - protein search, using sw model

Run On: May 12, 2003, 09:41:20 ; Search time 20 Seconds  
(without alignments)  
823.630 Million cell updates/sec

Title: US-09-995-493-52

Perfect score: 957

Sequence: 1 QHNGVLGPYIGKSLTLKLP.....ADYTTKATANLYGLNLNRYF 179

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	133.5	13.9	456	10	US-09-997-664-92
2	83.5	8.7	261	10	US-09-765-272-152
3	81.5	8.5	452	10	US-09-815-242-5141
4	80	8.4	164	9	US-09-854-133-203
5	80	8.4	164	9	US-09-738-973-203
6	80	8.4	698	9	US-09-935-642-6
7	78	8.2	367	10	US-09-881-752A-188
8	76	7.9	878	9	US-10-254-534-2
9	74	7.7	282	12	US-10-062-994-6
10	74	7.7	282	12	US-10-062-994-6
11	73.5	7.7	1698	9	US-10-080-943-4
12	72.5	7.6	406	10	US-09-925-297-517
13	72	7.5	1434	9	US-10-080-505-9
14	70.5	7.4	375	9	US-09-984-130-38
15	70.5	7.4	383	9	US-09-905-291A-261
16	70.5	7.4	383	9	US-09-902-853-261
17	70.5	7.4	383	9	US-09-907-824-261
18	70.5	7.4	383	9	US-09-907-841-261
19	70.5	7.4	383	9	US-09-904-011-261

20	70.5	7.4	383	9	US-10-174-590-50	Sequence 50, Appl
21	70.5	7.4	383	9	US-10-176-758-50	Sequence 50, Appl
22	70.5	7.4	383	9	US-10-175-737-50	Sequence 50, Appl
23	70.5	7.4	383	9	US-09-906-742-261	Sequence 261, App
24	70.5	7.4	383	9	US-10-173-706-50	Sequence 50, Appl
25	70.5	7.4	383	9	US-10-175-738-50	Sequence 50, Appl
26	70.5	7.4	383	9	US-10-175-752-50	Sequence 50, Appl
27	70.5	7.4	383	9	US-10-176-482-50	Sequence 50, Appl
28	70.5	7.4	383	9	US-10-176-757-50	Sequence 50, Appl
29	70.5	7.4	383	9	US-10-176-913-50	Sequence 50, Appl
30	70.5	7.4	383	9	US-10-180-552-50	Sequence 50, Appl
31	70.5	7.4	383	9	US-10-180-557-50	Sequence 50, Appl
32	70.5	7.4	383	9	US-09-906-838-261	Sequence 261, App
33	70.5	7.4	383	9	US-09-907-613-261	Sequence 261, App
34	70.5	7.4	383	9	US-09-907-942-261	Sequence 261, App
35	70.5	7.4	383	9	US-10-173-700-50	Sequence 50, Appl
36	70.5	7.4	383	9	US-10-174-572-50	Sequence 50, Appl
37	70.5	7.4	383	9	US-10-174-579-50	Sequence 50, Appl
38	70.5	7.4	383	9	US-10-174-582-50	Sequence 50, Appl
39	70.5	7.4	383	9	US-10-174-588-50	Sequence 50, Appl
40	70.5	7.4	383	9	US-10-175-739-50	Sequence 50, Appl
41	70.5	7.4	383	9	US-10-175-740-50	Sequence 50, Appl
42	70.5	7.4	383	9	US-10-175-743-50	Sequence 50, Appl
43	70.5	7.4	383	9	US-10-176-488-50	Sequence 50, Appl
44	70.5	7.4	383	9	US-10-176-492-50	Sequence 50, Appl
45	70.5	7.4	383	9	US-10-176-747-50	Sequence 50, Appl

## ALIGNMENTS

RESULT 1  
US-09-997-664-92  
; Sequence '92, Application US/09997664  
; Patent No. US20020151003A1  
; GENERAL INFORMATION:  
; APPLICANT: Ben-Bassat, Arie  
; APPLICANT: Cattermole, Monica  
; APPLICANT: Gatenby, Anthony A.  
; APPLICANT: Gibson, Katherine J.  
; APPLICANT: Ramos-Gonzalez, Isabel  
; APPLICANT: Ramos, Juan  
; APPLICANT: Sarlasiani, Sima  
; TITLE OF INVENTION: Method for the Production of p-Hydroxybenzoate in Species of  
; FILE REFERENCE: BC1018 US CIP  
; CURRENT FILING DATE: 2001-11-28  
; PRIOR FILING DATE: 2000-06-01  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 92  
; LENGTH: 456  
; TYPE: PRT  
; ORGANISM: Pseudomonas mendocina KR-1  
US-09-997-664-92

Query Match					
Best Local Similarity 26.7%, Pred. No. 1.9e-05; Length 456;					
Matches 40; Conservative 26; Mismatches 75; Indels 9; Gaps 4;					
QY	13	GSILTK---	LPAYWELSGFHQLTDQWAIHYSYKYTWSR-FKELRGKYDQSGCYEAFKK 68		
Db	291	GDIRVKDFEMPASLTGFAHQENRNLVAADYKRVVWSDMEDISVDFKSGGDIELP 350			
QY	69	BEYKNSRFAIGTYSNLDALTLAGLAYDKAASKTHLSAS-IPDTRMWYSGATYKFT 127			
Db	351	HNQDITVASICTAYRVNDKLTLAGYSYAOQALDSRLILPVIPAYLKKHVSUGDSYSD 410			
QY	128	P-----NLSDVDVGFALHURGKKKHVFETQNIK 153			
Db	411	KKSKLNLAIISFGLKESLNTPSYLSGTETLK 440			





```

; SEQ ID NO 6
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Anaplasma marginale
US-10-062-994-6

Query Match      7.7%; Score 74; DB 12; Length 282;
Best Local Similarity 24.5%; Pred. No. 11;
Matches 49; Conservative 27; Mismatches 86; Indels 38; Gaps 11;

QY 14 SLTLKLPAYWELSGF-----HQLTD-QWAIHYSY-----KYTEWSRFEKELRGKYQDGS 60
Db 84 TIDVSPANFSGKSYTFAFSKNLITSDGAVGYSLGARVELEASYYRREATLADGOYAKS 143
QY 61 GYE---AFTKKEEYKDNSRFAIGTTYSLNDALTLAGLAYDKAASKTHLS-----AS 109
Db 144 GAESLAAITRDANITETNYFVVKIDBITWTSVNLNG--CYDVLHTDLPVSPYVCAGIGAS 201
QY 110 IPD-----TDRMWY--SIGATYKFTPNLS--VDVGFAH--LRGKKKHVFETQNIKLLLVE 159
Db 202 FVDISKQVTTKLAYRGKVGISYQFTPEISLVAGGFVHGLFDESYKDIPAHNSVK--FSGE 259
QY 160 ADYTTKATANLYGLNLNRYF 179
Db 260 AKASVKAHIADYGFNLGARF 279

RESULT 10
US-10-062-994-6
; Sequence 6, Application US/10062994
; Patent No. US20020132789A1
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Bowle, Michael V.
; APPLICANT: Burrige, Michael J.
; APPLICANT: Mahan, Suman M.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Kurangirwa, Fred R.
; APPLICANT: Moreland, Annie L.
; APPLICANT: Simbi, Bigboy H.
; APPLICANT: Whitmire, William M.
; APPLICANT: Allemen, Arthur R.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and
; FILE REFERENCE: UF-167XC3
; CURRENT APPLICATION NUMBER: US/10/062,994
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/533,662
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 09/337,827
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 08/953,326
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 08/733,230
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Anaplasma marginale
US-10-062-994-6

Query Match      7.7%; Score 74; DB 12; Length 282;
Best Local Similarity 24.5%; Pred. No. 11;
Matches 49; Conservative 27; Mismatches 86; Indels 38; Gaps 11;

QY 14 SLTLKLPAYWELSGF-----HQLTD-QWAIHYSY-----KYTEWSRFEKELRGKYQDGS 60
Db 84 TIDVSPANFSGKSYTFAFSKNLITSDGAVGYSLGARVELEASYYRREATLADGOYAKS 143
QY 61 GYE---AFTKKEEYKDNSRFAIGTTYSLNDALTLAGLAYDKAASKTHLS-----AS 109
Db 144 GAESLAAITRDANITETNYFVVKIDBITWTSVNLNG--CYDVLHTDLPVSPYVCAGIGAS 201
QY 110 IPD-----TDRMWY--SIGATYKFTPNLS--VDVGFAH--LRGKKKHVFETQNIKLLLVE 159
Db 202 FVDISKQVTTKLAYRGKVGISYQFTPEISLVAGGFVHGLFDESYKDIPAHNSVK--FSGE 259
QY 160 ADYTTKATANLYGLNLNRYF 179
Db 260 AKASVKAHIADYGFNLGARF 279

RESULT 9
US-10-062-994-6
; Sequence 6, Application US/10062994
; Patent No. US2001001661A1
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Bowle, Michael V.
; APPLICANT: Burrige, Michael J.
; APPLICANT: Mahan, Suman M.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Kurangirwa, Fred R.
; APPLICANT: Moreland, Annie L.
; APPLICANT: Simbi, Bigboy H.
; APPLICANT: Whitmire, William M.
; APPLICANT: Allemen, Arthur R.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and
; FILE REFERENCE: UF-167XC3
; CURRENT APPLICATION NUMBER: US/10/062,994
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/533,662
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 09/337,827
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 08/953,326
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 08/733,230
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 878
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism:beII gene (branching enzyme II)
US-10-254-534-2

Query Match      7.9%; Score 76; DB 9; Length 878;
Best Local Similarity 34.8%; Pred. No. 28;
Matches 16; Conservative 10; Mismatches 18; Indels 2; Gaps 1;

QY 38 HVSRYKTEWSRFEKELRGKYQDGSYEATKKEEYKDNSRFAIGTTY 83
Db 193 HLDYRSQYKLRLEAIDYE--GGLEAFSGYKMGFTRSATGITY 236

RESULT 8
US-10-254-534-2
; Sequence 2, Application US/10254534
; Publication No. US20030046730A1
; GENERAL INFORMATION:
; APPLICANT: EK, Bo
; APPLICANT: KHOSNOODI, Jamshid
; APPLICANT: LARSSON, Clas-Tomas
; APPLICANT: LARSSON, Hakan
; APPLICANT: RASK, Lars
; TITLE OF INVENTION: STARCH BRANCHING ENZYME II OF POTATO
; FILE REFERENCE: 003300-486
; CURRENT APPLICATION NUMBER: US/10/254,534
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US/09/087,277
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: PCT/SE96/01558
; PRIOR FILING DATE: 1996-11-28
; PRIOR APPLICATION NUMBER: SE 9504272-7
; PRIOR FILING DATE: 1995-11-29
; PRIOR APPLICATION NUMBER: SE 9601506-0
; PRIOR FILING DATE: 1996-04-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 878
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism:beII gene (branching enzyme II)
US-10-254-534-2
```

```

; SEQ ID NO 6
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Anaplasma marginale
US-10-062-994-6

Query Match      7.7%; Score 74; DB 12; Length 282;
Best Local Similarity 24.5%; Pred. No. 11;
Matches 49; Conservative 27; Mismatches 86; Indels 38; Gaps 11;

QY 14 SLTLKLPAYWELSGF-----HQLTD-QWAIHYSY-----KYTEWSRFEKELRGKYQDGS 60
Db 84 TIDVSPANFSGKSYTFAFSKNLITSDGAVGYSLGARVELEASYYRREATLADGOYAKS 143
QY 61 GYE---AFTKKEEYKDNSRFAIGTTYSLNDALTLAGLAYDKAASKTHLS-----AS 109
Db 144 GAESLAAITRDANITETNYFVVKIDBITWTSVNLNG--CYDVLHTDLPVSPYVCAGIGAS 201
QY 110 IPD-----TDRMWY--SIGATYKFTPNLS--VDVGFAH--LRGKKKHVFETQNIKLLLVE 159
Db 202 FVDISKQVTTKLAYRGKVGISYQFTPEISLVAGGFVHGLFDESYKDIPAHNSVK--FSGE 259
QY 160 ADYTTKATANLYGLNLNRYF 179
Db 260 AKASVKAHIADYGFNLGARF 279

RESULT 10
US-10-062-994-6
; Sequence 6, Application US/10062994
; Patent No. US20020132789A1
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Bowle, Michael V.
; APPLICANT: Burrige, Michael J.
; APPLICANT: Mahan, Suman M.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Kurangirwa, Fred R.
; APPLICANT: Moreland, Annie L.
; APPLICANT: Simbi, Bigboy H.
; APPLICANT: Whitmire, William M.
; APPLICANT: Allemen, Arthur R.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and
; FILE REFERENCE: UF-167XC3
; CURRENT APPLICATION NUMBER: US/10/062,994
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/533,662
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 09/337,827
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 08/953,326
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 08/733,230
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Anaplasma marginale
US-10-062-994-6

Query Match      7.7%; Score 74; DB 12; Length 282;
Best Local Similarity 24.5%; Pred. No. 11;
Matches 49; Conservative 27; Mismatches 86; Indels 38; Gaps 11;

QY 14 SLTLKLPAYWELSGF-----HQLTD-QWAIHYSY-----KYTEWSRFEKELRGKYQDGS 60
Db 84 TIDVSPANFSGKSYTFAFSKNLITSDGAVGYSLGARVELEASYYRREATLADGOYAKS 143
QY 61 GYE---AFTKKEEYKDNSRFAIGTTYSLNDALTLAGLAYDKAASKTHLS-----AS 109
Db 144 GAESLAAITRDANITETNYFVVKIDBITWTSVNLNG--CYDVLHTDLPVSPYVCAGIGAS 201
QY 110 IPD-----TDRMWY--SIGATYKFTPNLS--VDVGFAH--LRGKKKHVFETQNIKLLLVE 159
Db 202 FVDISKQVTTKLAYRGKVGISYQFTPEISLVAGGFVHGLFDESYKDIPAHNSVK--FSGE 259
QY 160 ADYTTKATANLYGLNLNRYF 179
Db 260 AKASVKAHIADYGFNLGARF 279

RESULT 9
US-10-062-994-6
; Sequence 6, Application US/10062994
; Patent No. US2001001661A1
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Bowle, Michael V.
; APPLICANT: Burrige, Michael J.
; APPLICANT: Mahan, Suman M.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Kurangirwa, Fred R.
; APPLICANT: Moreland, Annie L.
; APPLICANT: Simbi, Bigboy H.
; APPLICANT: Whitmire, William M.
; APPLICANT: Allemen, Arthur R.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and
; FILE REFERENCE: UF-167XC3
; CURRENT APPLICATION NUMBER: US/10/062,994
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/533,662
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 09/337,827
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 08/953,326
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 08/733,230
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 878
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism:beII gene (branching enzyme II)
US-10-254-534-2

Query Match      7.9%; Score 76; DB 9; Length 878;
Best Local Similarity 34.8%; Pred. No. 28;
Matches 16; Conservative 10; Mismatches 18; Indels 2; Gaps 1;

QY 38 HVSRYKTEWSRFEKELRGKYQDGSYEATKKEEYKDNSRFAIGTTY 83
Db 193 HLDYRSQYKLRLEAIDYE--GGLEAFSGYKMGFTRSATGITY 236

RESULT 8
US-10-254-534-2
; Sequence 2, Application US/10254534
; Publication No. US20030046730A1
; GENERAL INFORMATION:
; APPLICANT: EK, Bo
; APPLICANT: KHOSNOODI, Jamshid
; APPLICANT: LARSSON, Clas-Tomas
; APPLICANT: LARSSON, Hakan
; APPLICANT: RASK, Lars
; TITLE OF INVENTION: STARCH BRANCHING ENZYME II OF POTATO
; FILE REFERENCE: 003300-486
; CURRENT APPLICATION NUMBER: US/10/254,534
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US/09/087,277
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: PCT/SE96/01558
; PRIOR FILING DATE: 1996-11-28
; PRIOR APPLICATION NUMBER: SE 9504272-7
; PRIOR FILING DATE: 1995-11-29
; PRIOR APPLICATION NUMBER: SE 9601506-0
; PRIOR FILING DATE: 1996-04-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 878
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism:beII gene (branching enzyme II)
US-10-254-534-2
```

Db 144 GAESLAAITRDANITETNVFVKIDEITNTSVMLNG--CYDVLHTLDPVSPYVCAGIGAS 201  
QY 110 IPD-----TDRMY--SIGATYKFTPNLS-VDVGFHFAH--LRGKKKHFEVETONIKGLLLVE 159  
Db 202 FVDISKQVTTKLAIRKGVGISQFTPEISLVAGGFYHGLFDESKDIPAHNSVK--FSGE 259  
QY 160 ADYTTKATANLYGLNLNRYF 179  
Db 260 AKASVRAHIADYGFNLGARF 279  
RESULT 11  
US-10-080-943-4  
; Sequence 4, Application US/10080943  
; Publication No. US20030073236A1  
; GENERAL INFORMATION:  
; APPLICANT: Field, Loren J.  
; APPLICANT: Tsai, Shih-Chong  
; TITLE OF INVENTION: P193 PROTEINS AND NUCLEIC ACIDS, AND USES THEREOF  
; FILE REFERENCE: IU99-PCT  
; CURRENT APPLICATION NUMBER: US/10/080,943  
; PRIOR FILING DATE: 2002-10-09  
; PRIOR APPLICATION NUMBER: 60/150,266  
; PRIOR FILING DATE: 1999-08-23  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 1698  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-080-943-4  
Query Match 7.7%; Score 73.5; DB 9; Length 1698;  
Best Local Similarity 23.9%; Pred. No. 1.2e+02;  
Matches 28; Conservative 17; Mismatches 47; Indels 25; Gaps 3;  
QY 64 AFTKKEYKDNSRFAITGTTYSLNDALTLRAGLAYDKAASKTHLSA-----SIPD 112  
Db 350 SFRSRFRPRSEFASGNTYALYVRTLOPGMRVRLDDYEISAGDEGEFROSGNGVPP 409  
QY 113 TDRMYSIGATYKFTPNLSVDVGFHRLGKKKHFEVETONIKGLLLVEADYTTKATAN 169  
Db 410 VQVFWESTGRTYVHHMLBILGF-----EEDIED--MVEADEYQGAVAS 452  
RESULT 12  
US-09-925-517  
; Sequence 517, Application US/09925297  
; Patent No. US20020081659A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA105  
; CURRENT APPLICATION NUMBER: US/09/925,297  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 517  
; LENGTH: 406  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-517  
Query Match 7.6%; Score 72.5; DB 10; Length 406;  
Best Local Similarity 21.7%; Pred. No. 24;  
Matches 45; Conservative 25; Mismatches 66; Indels 71; Gaps 12;  
QY 18 KLPAYWELSGFHLTDQWATHYSKYKTEMRSFKELRGKYQDGSYEAFTHK--EYKDNS 75

Db 173 KRPAIWDITGIIH--SREWVTQAS---GVW-----FAKKKITQDYCQDA 209  
QY 76 REA--IGTYSLNDALTLRAGLAYDKAA-----SKTHLSASI---PDTDRMYS----- 119  
Db 210 AFTAILDTLDIFLEIVTNPDPGFAFTHTNRMRKRTKTSHTAGSLCIGVDPRNWDAGFGLS 269  
QY 120 -----IGATY--KFTPNLSVD-----VGFHRLGKKKHFEVETONIKGLLLVEADYTTK-- 165  
Db 270 GASSNPCSEYTHGKFA-NSEVEVKSIQVDFVKDHGNIKAFISHSYSQLLMYPYGYKTEPV 328  
QY 166 -----ATANLYGLNLNRY 177  
Db 329 PQDELDQLSKAAAVTALASLYGTFKNY 355  
RESULT 13  
US-10-080-505-9  
; Sequence 9, Application US/10080505  
; Publication No. US20030073166A1  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme, Joseph W.  
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS  
; FILE REFERENCE: A-59941-1/RET/DCF/DHR  
; CURRENT APPLICATION NUMBER: US/10/080,505  
; CURRENT FILING DATE: 2002-02-22  
; PRIOR APPLICATION NUMBER: US 08/296,791  
; PRIOR FILING DATE: 1994-10-25  
; PRIOR APPLICATION NUMBER: US 09/839,996  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: Patent In version 3.1  
; SEQ ID NO 9  
; LENGTH: 1434  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (568)..(568)  
; OTHER INFORMATION: The 'Xaa' at location 568 stands for Ser, Gly, Arg, or Cys.  
; NAME/KEY: misc\_feature  
; LOCATION: (1702)..(1702)  
; OTHER INFORMATION: "n" at position 1702 can be any base.  
US-10-080-505-9  
Query Match 7.5%; Score 72; DB 9; Length 1434;  
Best Local Similarity 19.1%; Pred. No. 1.3e+02;  
Matches 34; Conservative 30; Mismatches 56; Indels 58; Gaps 9;  
QY 46 WSRFEKELRGKYQDGSYEAFTHKKEEYK-----DNSRFAITGTTYSLND----- 87  
Db 1189 WTNIADQKRY-DSDAFRAYQOKTNLRQIGVQKALDNGRIGAVFSHRSRSDNTEDEQVKNH 1247  
QY 88 -ALTLRAGLAY-----DKAASKTHLSASIPDTRMYSIGATY 124  
Db 1248 ATTTMMSGFAQYQWGDLOFGVNVGAGISASKMAEEQSKTHRKA-----INGVNASY 1300  
QY 125 KFTPNLSVDVGFHRLGKKKHFEVETONIKGLLLVEADYTTKATA-NLY--GLNLNRYF 179  
Db 1301 QFRIG---QLGIQPYLGVNRYFIERYNQS---EEVKVQTPSLAFNRYNAGIRVDYTF 1352  
RESULT 14  
US-09-984-130-38  
; Sequence 38, Application US/09984130  
; Publication No. US20030055231A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 12 Human Secreted Proteins  
; FILE REFERENCE: PF489p2  
; CURRENT APPLICATION NUMBER: US/09/984,130  
; CURRENT FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: 60/243,792  
; PRIOR FILING DATE: 2000-10-30

;; PRIOR APPLICATION NUMBER: 09/836,353  
;; PRIOR FILING DATE: 2001-04-18  
;; PRIOR APPLICATION NUMBER: 60/198,407  
;; PRIOR FILING DATE: 2000-04-19  
;; PRIOR APPLICATION NUMBER: PCT/US99/25031  
;; PRIOR FILING DATE: 1999-10-27  
;; PRIOR APPLICATION NUMBER: 60/105,971  
;; PRIOR FILING DATE: 1998-10-28  
;; NUMBER OF SEQ ID NOS: 149  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 38  
;; LENGTH: 375  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-984-130-38

Query Match 7.4%; Score 70.5; DB 9; Length 375;  
Best Local Similarity 23.8%; Pred. No. 35;  
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;  
Qy 7 GPYIGKSLTLKLPAYWELSGFHLTDQWATHYSKYKTEWSRKFELRGKY-----LSYETLVANGSRTTQVGIYILSSSGDGAQ 114  
Db 66 GPQCHKGT---PLPYEEAKQY-----LSYETLVANGSRTTQVGIYILSSSGDGAQ 114  
Qy 57 --QDGSYEAFPTKKEEYKDNSRFAI-GTTYSLN-----DALTLRAGLAYDKAAKSTHLSAS 109  
Db 115 HRDSSGSKSRKRQIYGYDSRFSIFGKDFLLNYPFSTSVKLSGTCTGTLVAEKHVLTA 174  
Qy 110 --IPDTRMYSICATY-KETPNLSV-----DVGFAHLRGKK 143  
Db 175 HCHID-----GKTVKGTQKLRVGLFKPKFDGGRGANDSTSAMPEQMKFQWIRVKR 226  
Qy 144 KHFEVTONIKGLLVEADYTTKATANLYGLNLN 177  
Db 227 THVP-----KGWI-----KGNANDIGMDYDY 247

RESULT 15  
US-09-905-291A-261  
; Sequence 261, Application US/09905291A  
; Patent No. US20020160374A1  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; APPLICANT: Ashkenazi, Avi  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Flivaroff, Ellen  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, A.  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth, J.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Mather, Jennie P.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William, I.  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; TITLE OF INVENTION: Acids Encoding the Same  
; FILE REFERENCE: 10466-14  
; CURRENT APPLICATION NUMBER: US/09/905,291A  
; CURRENT FILING DATE: 2001-07-12  
; PRIOR APPLICATION NUMBER: PCT/US00/04414

;; PRIOR FILING DATE: 2000-02-22  
;; PRIOR APPLICATION NUMBER: US 60/143,048  
;; PRIOR FILING DATE: 1999-07-07  
;; PRIOR APPLICATION NUMBER: US 60/145,698  
;; PRIOR FILING DATE: 1999-07-26  
;; PRIOR APPLICATION NUMBER: US 60/146,222  
;; PRIOR FILING DATE: 1999-07-28  
;; PRIOR APPLICATION NUMBER: PCT/US99/20594  
;; PRIOR FILING DATE: 1999-09-08  
;; PRIOR APPLICATION NUMBER: PCT/US99/20944  
;; PRIOR FILING DATE: 1999-09-13  
;; PRIOR APPLICATION NUMBER: PCT/US99/21090  
;; PRIOR FILING DATE: 1999-09-15  
;; PRIOR APPLICATION NUMBER: PCT/US99/21547  
;; PRIOR FILING DATE: 1999-09-15  
;; PRIOR APPLICATION NUMBER: PCT/US99/23089  
;; PRIOR FILING DATE: 1999-10-05  
;; PRIOR APPLICATION NUMBER: PCT/US99/28214  
;; PRIOR FILING DATE: 1999-11-29  
;; PRIOR APPLICATION NUMBER: PCT/US99/28313  
;; PRIOR FILING DATE: 1999-11-30  
;; PRIOR APPLICATION NUMBER: PCT/US99/28564  
;; PRIOR FILING DATE: 1999-12-02  
;; PRIOR APPLICATION NUMBER: PCT/US99/28565  
;; PRIOR FILING DATE: 1999-12-02  
;; PRIOR APPLICATION NUMBER: PCT/US99/30095  
;; PRIOR FILING DATE: 1999-12-16  
;; PRIOR APPLICATION NUMBER: PCT/US99/30911  
;; PRIOR FILING DATE: 1999-12-20  
;; PRIOR APPLICATION NUMBER: PCT/US99/30999  
;; PRIOR FILING DATE: 1999-12-20  
;; PRIOR APPLICATION NUMBER: PCT/US00/00219  
;; NUMBER OF SEQ ID NOS: 423  
;; SEQ ID NO 261  
;; LENGTH: 383  
;; TYPE: PRT  
;; ORGANISM: Homo Sapien  
US-09-905-291A-261  
Query Match 7.4%; Score 70.5; DB 9; Length 383;  
Best Local Similarity 23.8%; Pred. No. 36;  
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;  
Qy 7 GPYIGKSLTLKLPAYWELSGFHLTDQWATHYSKYKTEWSRKFELRGKY-----LSYETLVANGSRTTQVGIYILSSSGDGAQ 114  
Db 66 GPQCHKGT---PLPYEEAKQY-----LSYETLVANGSRTTQVGIYILSSSGDGAQ 114  
Qy 57 --QDGSYEAFPTKKEEYKDNSRFAI-GTTYSLN-----DALTLRAGLAYDKAAKSTHLSAS 109  
Db 115 HRDSSGSKSRKRQIYGYDSRFSIFGKDFLLNYPFSTSVKLSGTCTGTLVAEKHVLTA 174  
Qy 110 --IPDTRMYSICATY-KETPNLSV-----DVGFAHLRGKK 143  
Db 175 HCHID-----GKTVKGTQKLRVGLFKPKFDGGRGANDSTSAMPEQMKFQWIRVKR 226  
Qy 144 KHFEVTONIKGLLVEADYTTKATANLYGLNLN 177  
Db 227 THVP-----KGWI-----KGNANDIGMDYDY 247

RESULT 16  
US-09-902-853-261  
; Sequence 261, Application US/09902853  
; Publication No. US20020192659A1  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; APPLICANT: Ashkenazi, Avi  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Flivaroff, Ellen





```

RESULT 18
US/09-907-841-261
Sequence 261, Application US/09907841
Publication No. US20020198366A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Garber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Tra
FILE OF INVENTION: Acids Encoding
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/90
CURRENT FILING DATE: 2001-11-20

```

RESULT 19  
US - 09-904 - 011 - 261  
Sequence 261, Application US/09904011  
Publication No. US20030003530A1  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc.  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, A.  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth, J.  
APPLICANT: Kijavlin, Ivar J.  
APPLICANT: Mather, Jennie P.  
APPLICANT: Pan, James

APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William, I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: 10466-14  
CURRENT APPLICATION NUMBER: US/09/904,011  
PRIOR FILING DATE: 2001-07-11  
PRIOR APPLICATION NUMBER: 09/665,350  
PRIOR FILING DATE: 2000-09-18  
PRIOR APPLICATION NUMBER: PCT/US00/04414  
PRIOR FILING DATE: 2000-02-22  
PRIOR APPLICATION NUMBER: US 60/143,048  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: US 60/145,698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: US 60/146,222  
PRIOR FILING DATE: 1999-07-28  
PRIOR APPLICATION NUMBER: PCT/US99/20594  
PRIOR FILING DATE: 1999-09-08  
PRIOR APPLICATION NUMBER: PCT/US99/20944  
PRIOR FILING DATE: 1999-09-13  
PRIOR APPLICATION NUMBER: PCT/US99/21090  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/21547  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/23089  
PRIOR FILING DATE: 1999-10-05  
PRIOR APPLICATION NUMBER: PCT/US99/28214  
PRIOR FILING DATE: 1999-11-29  
PRIOR APPLICATION NUMBER: PCT/US99/28313  
PRIOR FILING DATE: 1999-11-30  
PRIOR APPLICATION NUMBER: PCT/US99/28564  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/28565  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/30095  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: PCT/US99/30911  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US99/30999  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US00/00219  
PRIOR FILING DATE: 2000-01-05  
NUMBER OF SEQ ID NOS: 423  
SEQ ID NO 261  
LENGTH: 383  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-09-904-011-261

Query Match 7.4%; Score 70.5; DB 9; Length 383;  
Best Local Similarity 23.8%; Pred. No. 36;  
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;  
QY 7 GPYIGKSLTLKLPAYWELSGFHLTDQWAIHYSYKYTEMSRKEKLGKY-----56  
Db 66 GPOCHKGT---PLPTYEAKQY-----LSYETLYANGSRTEQTQVGIYILSSSGDGAQ 114  
QY 57 --QDGSYGAFTKKEEYKONSRAFI-GTYYSLN-----DALTLRAGLAYDKAASKTHLSAS 109  
Db 115 HRDSSGSKSRKRIQYGYDSRFSIFGKDFLLNYPSTSVKLGSTGCTGTLVAEKHVLTA 174  
QY 110 --IPDTRMWSYGATY-KFTPNLSV-----DVGFAHLRGKK 143  
Db 175 HCIHD-----GKTYVKGTKLRVGLPKPKDGGRGANDSTSAMPEQMKFQWIRVKR 226  
QY 144 KHFVETQNIKGLLLVEADYTTKATANLYGLNLNY 177  
Db 227 THVP-----KGWI-----KGNANDIGMDYDY 247  
US-09-904-011-261

RESULT 20  
US-10-174-590-50  
Sequence 50, Application US/10174590  
Publication No. US20030008352A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430RIC42  
CURRENT APPLICATION NUMBER: US/10/174,590  
CURRENT FILING DATE: 2002-06-18  
Prior application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 612  
SEQ ID NO 50  
LENGTH: 383  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-174-590-50  
Query Match 7.4%; Score 70.5; DB 9; Length 383;  
Best Local Similarity 23.8%; Pred. No. 36;  
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;  
QY 7 GPYIGKSLTLKLPAYWELSGFHLTDQWAIHYSYKYTEMSRKEKLGKY-----56  
Db 66 GPOCHKGT---PLPTYEAKQY-----LSYETLYANGSRTEQTQVGIYILSSSGDGAQ 114  
QY 57 --QDGSYGAFTKKEEYKONSRAFI-GTYYSLN-----DALTLRAGLAYDKAASKTHLSAS 109  
Db 115 HRDSSGSKSRKRIQYGYDSRFSIFGKDFLLNYPSTSVKLGSTGCTGTLVAEKHVLTA 174  
QY 110 --IPDTRMWSYGATY-KFTPNLSV-----DVGFAHLRGKK 143  
Db 175 HCIHD-----GKTYVKGTKLRVGLPKPKDGGRGANDSTSAMPEQMKFQWIRVKR 226  
QY 144 KHFVETQNIKGLLLVEADYTTKATANLYGLNLNY 177  
Db 227 THVP-----KGWI-----KGNANDIGMDYDY 247

RESULT 21  
US-10-176-758-50  
Sequence 50, Application US/10176758  
Publication No. US20030008353A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430RIC104  
CURRENT APPLICATION NUMBER: US/10/176,758  
CURRENT FILING DATE: 2002-06-21

; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 50  
; LENGTH: 383  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-176-758-50

Query Match 7.4%; Score 70.5; DB 9; Length 383;  
Best Local Similarity 23.8%; Pred. No. 36;  
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;  
  
Qy 7 GPTGKSLTLKLPAYWELSGHQLTDQWATHYSKYKTEWRFKELRGKY----- 56  
Db 66 GPOCHKGT---PLPTYEAKQY-----LSYETLYANGSRRTETQVGIIYLSSSGDAQ 114  
Qy 57 --QDGSYEATFKREYKDNRSFAT-GTYSLN-----DALTLRAGLAYDKAASKTHLSAS 109  
Db 115 HRDSSGCKSRKRQIYGYDSRFSIEGKDFLLNYPFSTSVKLSGCTGLVAEKHVLTA 174  
Qy 110 --IPDTRMWSIGATY-KFTPNLSV-----DVGFALHGRKK 143  
Db 175 HCIHD-----CKTYVKGTQKLRVGLPKPKFDGGRGANDSTSAMPEQMKFQWIRVKR 226  
Qy 144 KHVETONIKGLLLVEADYTTKATANLYGLNLNY 177  
Db 227 THVP-----KGWI-----KGNANDIGMDYDY 247

## RESULT 22

US-10-175-737-50  
; Sequence 50, Application US/10175737  
; Publication No. US20030013153A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Goddard, Paul J.  
; APPLICANT: Goddard, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3430R1C50  
; CURRENT APPLICATION NUMBER: US/10/175,737  
; CURRENT FILING DATE: 2002-06-19  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 50  
; LENGTH: 383  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-175-737-50

Query Match 7.4%; Score 70.5; DB 9; Length 383;  
Best Local Similarity 23.8%; Pred. No. 36;  
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;  
  
Qy 7 GPTGKSLTLKLPAYWELSGHQLTDQWATHYSKYKTEWRFKELRGKY----- 56  
Db 66 GPOCHKGT---PLPTYEAKQY-----LSYETLYANGSRRTETQVGIIYLSSSGDAQ 114  
Qy 57 --QDGSYEATFKREYKDNRSFAT-GTYSLN-----DALTLRAGLAYDKAASKTHLSAS 109  
Db 115 HRDSSGCKSRKRQIYGYDSRFSIEGKDFLLNYPFSTSVKLSGCTGLVAEKHVLTA 174  
Qy 110 --IPDTRMWSIGATY-KFTPNLSV-----DVGFALHGRKK 143  
Db 175 HCIHD-----CKTYVKGTQKLRVGLPKPKFDGGRGANDSTSAMPEQMKFQWIRVKR 226

Qy 144 KHVETONIKGLLLVEADYTTKATANLYGLNLNY 177  
Db 227 THVP-----KGWI-----KGNANDIGMDYDY 247  
  
RESULT 23  
US-09-906-742-261  
; Sequence 261, Application US/09906742  
; Publication No. US20030023054A1  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; APPLICANT: Ashkenazi, Avi  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Flivaroff, Ellen  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, A.  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth, J.  
; APPLICANT: Kijavlin, Ivar J.  
; APPLICANT: Mather, Jennie P.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; TITLE OF INVENTION: Acids Encoding the Same  
; FILE REFERENCE: 10466-14  
; CURRENT APPLICATION NUMBER: US/09/906,742  
; CURRENT FILING DATE: 2001-07-16  
; PRIOR APPLICATION NUMBER: 09/665,350  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: PCT/US00/04414  
; PRIOR FILING DATE: 2000-02-22  
; PRIOR APPLICATION NUMBER: US 60/143,048  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 60/145,698  
; PRIOR FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: US 60/146,222  
; PRIOR FILING DATE: 1999-07-28  
; PRIOR APPLICATION NUMBER: PCT/US99/20594  
; PRIOR FILING DATE: 1999-09-08  
; PRIOR APPLICATION NUMBER: PCT/US99/20944  
; PRIOR FILING DATE: 1999-09-13  
; PRIOR APPLICATION NUMBER: PCT/US99/21090  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: PCT/US99/21547  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: PCT/US99/23089  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: PCT/US99/28214  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: PCT/US99/28313  
; PRIOR FILING DATE: 1999-11-30  
; PRIOR APPLICATION NUMBER: PCT/US99/28564  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: PCT/US99/28565  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: PCT/US99/30095  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: PCT/US99/30911  
; PRIOR FILING DATE: 1999-12-20

```

QY    110 --IPDTRDMWYSIGATY-KFTPNLSV-----DVGF AHLRGKK 143
Db    175 HCLHD-----GKTVVKGTOKLRVGFELKPKFKDGGRANDSTSAMPEQMKEFQWRVRKR 226

QY    144 KHEVETQNIKGLLLVLEADYYTTKATANLYGLNLNY 177
Db    227 THVP-----KGWI-----KGNANDIGMDYDY 247

RESULT 25
US-10-175-738-50
; Sequence 50, Application US/10175738
; Publication No. US2003002294A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Godowski, Austin L.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430RIC45
; CURRENT APPLICATION NUMBER: US/10/175,738
; CURRENT FILING DATE: 2002-06-19
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 50
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-738-50

Query Match          7.4%; Score 70.5; DB 9; Length 383;
Best Local Similarity 23.8%; Pred. No. 36;
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps

QY    7 GPYIGRGSULTKLPAWELSGFHQLTDWAIIHYSKYIETWSRFKELRGKY-----56
Db    66 GPQCHGT---PLPTEAAKQY-----LSYETLYANGSRTEQTGVGIYILSSSGDGAQ 114

QY    57 --ODGSGYEAFTKKEEYKDNSRPAL-GTTYSLN----DALTLRAGLAYDKAKSKTHLSAS 109
Db    115 HRDSGSGSRRRKQRQIYGVDRSFEIGKDFLLNYPFSTSVKLTGCTGTLVAEKHVLTAA 174

QY    110 --IPDTRDMWYSIGATY-KFTPNLSV-----DVGF AHLRGKK 143
Db    175 HCLHD-----GKTVVKGTOKLRVGFELKPKFKDGGRANDSTSAMPEQMKEFQWRVRKR 226

QY    144 KHEVETQNIKGLLLVLEADYYTTKATANLYGLNLNY 177
Db    227 THVP-----KGWI-----KGNANDIGMDYDY 247

```

US-10-175-752-50  
; Sequence 50, Application US/10175752  
; Publication No. US2003002295A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.

```

; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C60
; CURRENT APPLICATION NUMBER: US/10/175,752
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 50
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Saplen
US-10-175-752-50

Query Match      7.4%; Score 70.5; DB 9; Length 383;
Best Local Similarity 23.8%; Pred. No. 36;
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;

Qy 7 GPYIGKSLTLKPAYWELSGFHOLTOWAIIHYSKYKTEWSRFKELRGKY----- 56
Db 66 GPQCHKGT---PLPTYEAKQY-----LSYETLVANGSRRTETQVGIVILSSSGDGAQ 114

Qy 57 --QDGSYEAFTKKEEYKDNRSFAI-GTTYSLN-----DALTLRAGLAYDKAASKTHLSAS 109
Db 115 HRDSSGSKSRKQIYGYDSRFSIFGKDFLLNYPFSTSVKLSGCTGLTVAEKHVLTA 174

Qy 110 --IPDTRMWSIGATY-KFTPNLSV-----DVGFALHGRGKK 143
Db 175 HCIHD-----GKTVYKGTQKLRVGLFKPKFDGGRGANDSTSAMPEQMKFQWIRVKR 226

Qy 144 KHFEVTONIKGLLVEADYTTKATANLYGLNLNY 177
Db 227 THVP-----KGWI-----KGNANDIGMDYD 247

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C86
; CURRENT APPLICATION NUMBER: US/10/176,757
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 50
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Saplen
US-10-176-757-50
```

```

RESULT 27
US-10-176-482-50
; Sequence 50, Application US/10176482
; Publication No. US20030022296A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C70
; CURRENT APPLICATION NUMBER: US/10/176,482
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 50
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Saplen
US-10-176-482-50

Query Match      7.4%; Score 70.5; DB 9; Length 383;
Best Local Similarity 23.8%; Pred. No. 36;
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;

Qy 7 GPYIGKSLTLKPAYWELSGFHOLTOWAIIHYSKYKTEWSRFKELRGKY----- 56
Db 66 GPQCHKGT---PLPTYEAKQY-----LSYETLVANGSRRTETQVGIVILSSSGDGAQ 114
```

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C70
; CURRENT APPLICATION NUMBER: US/10/176,482
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 50
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Saplen
US-10-176-482-50
```

```

Qy 57 --QDGSYEAFTKKEEYKDNRSFAI-GTTYSLN-----DALTLRAGLAYDKAASKTHLSAS 109
Db 115 HRDSSGSKSRKQIYGYDSRFSIFGKDFLLNYPFSTSVKLSGCTGLTVAEKHVLTA 174

Qy 110 --IPDTRMWSIGATY-KFTPNLSV-----DVGFALHGRGKK 143
Db 175 HCIHD-----GKTVYKGTQKLRVGLFKPKFDGGRGANDSTSAMPEQMKFQWIRVKR 226

Qy 144 KHFEVTONIKGLLVEADYTTKATANLYGLNLNY 177
Db 227 THVP-----KGWI-----KGNANDIGMDYD 247

RESULT 28
US-10-176-757-50
; Sequence 50, Application US/10176757
; Publication No. US20030022297A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C86
; CURRENT APPLICATION NUMBER: US/10/176,757
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 50
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Saplen
US-10-176-757-50

Query Match      7.4%; Score 70.5; DB 9; Length 383;
Best Local Similarity 23.8%; Pred. No. 36;
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;

Qy 7 GPYIGKSLTLKPAYWELSGFHOLTOWAIIHYSKYKTEWSRFKELRGKY----- 56
Db 66 GPQCHKGT---PLPTYEAKQY-----LSYETLVANGSRRTETQVGIVILSSSGDGAQ 114

Qy 57 --QDGSYEAFTKKEEYKDNRSFAI-GTTYSLN-----DALTLRAGLAYDKAASKTHLSAS 109
Db 115 HRDSSGSKSRKQIYGYDSRFSIFGKDFLLNYPFSTSVKLSGCTGLTVAEKHVLTA 174

Qy 110 --IPDTRMWSIGATY-KFTPNLSV-----DVGFALHGRGKK 143
Db 175 HCIHD-----GKTVYKGTQKLRVGLFKPKFDGGRGANDSTSAMPEQMKFQWIRVKR 226

Qy 144 KHFEVTONIKGLLVEADYTTKATANLYGLNLNY 177
Db 227 THVP-----KGWI-----KGNANDIGMDYD 247

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C86
; CURRENT APPLICATION NUMBER: US/10/176,757
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 50
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Saplen
US-10-176-757-50
```

```

Query Match      7.4%; Score 70.5; DB 9; Length 383;
Best Local Similarity 23.8%; Pred. No. 36;
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;

Qy 7 GPYIGKSLTLKPAYWELSGFHOLTOWAIIHYSKYKTEWSRFKELRGKY----- 56
Db 66 GPQCHKGT---PLPTYEAKQY-----LSYETLVANGSRRTETQVGIVILSSSGDGAQ 114

Qy 57 --QDGSYEAFTKKEEYKDNRSFAI-GTTYSLN-----DALTLRAGLAYDKAASKTHLSAS 109
Db 115 HRDSSGSKSRKQIYGYDSRFSIFGKDFLLNYPFSTSVKLSGCTGLTVAEKHVLTA 174

Qy 110 --IPDTRMWSIGATY-KFTPNLSV-----DVGFALHGRGKK 143
Db 175 HCIHD-----GKTVYKGTQKLRVGLFKPKFDGGRGANDSTSAMPEQMKFQWIRVKR 226

Qy 144 KHFEVTONIKGLLVEADYTTKATANLYGLNLNY 177
Db 227 THVP-----KGWI-----KGNANDIGMDYD 247

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C86
; CURRENT APPLICATION NUMBER: US/10/176,757
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 50
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Saplen
US-10-176-757-50
```

```

RESULT 29
US-10-176-913-50
; Sequence 50, Application US/10176913
; Publication No. US20030022298A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C70
; CURRENT APPLICATION NUMBER: US/10/176,913
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 50
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Saplen
US-10-176-913-50
```

APPLICANT: Gurney,Austin L.  
APPLICANT: Pan,James  
APPLICANT: Smith,Victoria  
APPLICANT: Watanabe,Colin K.  
APPLICANT: Wood,William I.  
APPLICANT: Zhang,Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430RIC66  
CURRENT APPLICATION NUMBER: US/10/176,913  
Prior Filing DATE: 2002-06-20  
NUMBER OF SEQ ID NOS: 612  
SEQ ID NO 50  
LENGTH: 383  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-176-913-50

Query Match 7.4%; Score 70.5; DB 9; Length 383;  
Best Local Similarity 23.8%; Pred. No. 36;  
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;

QY 7 GPYIGKGSUTLKLPAWELSGFHLTDQWAIHYSYKYTEWSRPFKELRGY-----56  
Db 66 GPQCHKGT---PLPTVEEAKQY-----LSYETLYANGSRTEQTQVGIYILSSSGDGAQ 114  
QY 57 --QDGSGEAFYKKEEYKONSFPAL-GTYYSLN---DALTLRAGLAYDKAASKTHLSAS 109  
Db 115 HRDSSGSGRRKRQIYGYDSRFSIFGKDFLLNYPSTSVKLTGCTGLTVAEKHVLTA 174  
QY 110 --IPDTRMWSYGATY-KFTPNLSV-----DVGFAHLRGKK 143  
Db 175 HCHID-----GKTYVKGTKLRVGFLEKPKFKDGGGRANDSTSAMPEOMKQFOWIRVKR 226  
QY 144 KHFEVTONIKGLLLVEADYTTATANLYGLNLNY 177  
Db 227 THVP-----KGWI-----KGNANDIGMDYDY 247

RESULT 30  
US-10-180-552-50  
Sequence 50, Application US/10180552  
Publication NO. US20030022300A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430RIC153  
CURRENT APPLICATION NUMBER: US/10/180,552  
Prior Filing DATE: 2002-06-25  
NUMBER OF SEQ ID NOS: 612  
SEQ ID NO 50  
LENGTH: 383  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-180-552-50

Query Match 7.4%; Score 70.5; DB 9; Length 383;  
Best Local Similarity 23.8%; Pred. No. 36;  
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;

QY 7 GPYIGKGSUTLKLPAWELSGFHLTDQWAIHYSYKYTEWSRPFKELRGY-----56  
Db 66 GPQCHKGT---PLPTVEEAKQY-----LSYETLYANGSRTEQTQVGIYILSSSGDGAQ 114  
QY 57 --QDGSGEAFYKKEEYKONSFPAL-GTYYSLN---DALTLRAGLAYDKAASKTHLSAS 109  
Db 115 HRDSSGSGRRKRQIYGYDSRFSIFGKDFLLNYPSTSVKLTGCTGLTVAEKHVLTA 174  
QY 110 --IPDTRMWSYGATY-KFTPNLSV-----DVGFAHLRGKK 143  
Db 175 HCHID-----GKTYVKGTKLRVGFLEKPKFKDGGGRANDSTSAMPEOMKQFOWIRVKR 226  
QY 144 KHFEVTONIKGLLLVEADYTTATANLYGLNLNY 177  
Db 227 THVP-----KGWI-----KGNANDIGMDYDY 247

RESULT 31  
US-10-180-557-50  
Sequence 50, Application US/10180557  
Publication NO. US20030022301A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430RIC147  
CURRENT APPLICATION NUMBER: US/10/180,557  
Prior Filing DATE: 2002-06-25  
NUMBER OF SEQ ID NOS: 612  
SEQ ID NO 50  
LENGTH: 383  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-180-557-50

Query Match 7.4%; Score 70.5; DB 9; Length 383;  
Best Local Similarity 23.8%; Pred. No. 36;  
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;

QY 7 GPYIGKGSUTLKLPAWELSGFHLTDQWAIHYSYKYTEWSRPFKELRGY-----56  
Db 66 GPQCHKGT---PLPTVEEAKQY-----LSYETLYANGSRTEQTQVGIYILSSSGDGAQ 114  
QY 57 --QDGSGEAFYKKEEYKONSFPAL-GTYYSLN---DALTLRAGLAYDKAASKTHLSAS 109  
Db 115 HRDSSGSGRRKRQIYGYDSRFSIFGKDFLLNYPSTSVKLTGCTGLTVAEKHVLTA 174  
QY 110 --IPDTRMWSYGATY-KFTPNLSV-----DVGFAHLRGKK 143  
Db 175 HCHID-----GKTYVKGTKLRVGFLEKPKFKDGGGRANDSTSAMPEOMKQFOWIRVKR 226  
QY 144 KHFEVTONIKGLLLVEADYTTATANLYGLNLNY 177  
Db 227 THVP-----KGWI-----KGNANDIGMDYDY 247

RESULT 32  
US-09-906-838-261  
Sequence 261, Application US/09906838  
Publication NO. US20030027143A1  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc.

APPLICANT: Ashkenazi, Avi  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, A.  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth, J.  
APPLICANT: Kijavlin, Ivar J.  
APPLICANT: Mather, Jennie P.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William, I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
Acids Encoding the Same  
FILE REFERENCE: 10466-14  
CURRENT APPLICATION NUMBER: US/09/906,838  
CURRENT FILING DATE: 2001-07-16  
PRIOR APPLICATION NUMBER: 09/665,350  
PRIOR FILING DATE: 2000-09-18  
PRIOR APPLICATION NUMBER: PCT/US00/04414  
PRIOR FILING DATE: 2000-02-22  
PRIOR APPLICATION NUMBER: US 60/143,048  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: US 60/145,698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: US 60/146,222  
PRIOR FILING DATE: 1999-07-28  
PRIOR APPLICATION NUMBER: PCT/US99/20594  
PRIOR FILING DATE: 1999-09-08  
PRIOR APPLICATION NUMBER: PCT/US99/20944  
PRIOR FILING DATE: 1999-09-13  
PRIOR APPLICATION NUMBER: PCT/US99/21090  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/21547  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/23089  
PRIOR FILING DATE: 1999-10-05  
PRIOR APPLICATION NUMBER: PCT/US99/28214  
PRIOR FILING DATE: 1999-11-29  
PRIOR APPLICATION NUMBER: PCT/US99/28313  
PRIOR FILING DATE: 1999-11-30  
PRIOR APPLICATION NUMBER: PCT/US99/28564  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/28565  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/30095  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: PCT/US99/30911  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US99/30999  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US00/00219  
PRIOR FILING DATE: 2000-01-05  
NUMBER OF SEQ ID NOS: 423  
SEQ ID NO 261  
LENGTH: 383  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-09-906-838-261

Query Match 7.4%; Score 70.5; DB 9; Length 383;

Best Local Similarity 23.8%; Pred. No. 36;  
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;  
QY 7 GPYIGKSLTLKLPAYWELSCFHOLTOWAIHYSKYKTEWSRPFELRGKY-----56  
Db 66 GPQCHKGT---PLPTYBEAKQY-----LSYETLYANGSRRTETQVGIYILSSSGDGAQ 114  
QY 57 --QDGSVEAFTKKEEYKDNRSFAI-GTYSLN----DALTLRAGLAYDKAASKTHLSAS 109  
Db 115 HROSGSSCKSRKQIYGYDSRFSIFGKDFLLNYPFSTSVKLSGCTGTLVAEKHVLTA 174  
QY 110 --IPDTRMMYSIGATY-KFTPNLSV-----DVGFAHLRGKK 143  
Db 175 HCIHD-----GKTVYKGTQKLRVGLKPKFKDGGRGANDSTSAMPEQMKFQWIRVKR 226  
QY 144 KHFEVETQNIKGLLVEADYTTKATANLYGLNLNY 177  
Db 227 THVP-----KGWI-----KGNANDIGMDYD 247  
RESULT 33  
US-09-907-613-261  
Sequence 261, Application US/09907613  
Publication No. US20030027145A1.  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc.  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, A.  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth, J.  
APPLICANT: Kijavlin, Ivar J.  
APPLICANT: Mather, Jennie P.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William, I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
Acids Encoding the Same  
FILE REFERENCE: 10466-14  
CURRENT APPLICATION NUMBER: US/09/907,613  
CURRENT FILING DATE: 2001-07-17  
PRIOR APPLICATION NUMBER: PCT/US00/04414  
PRIOR FILING DATE: 2000-02-22  
PRIOR APPLICATION NUMBER: US 60/143,048  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: US 60/145,698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: US 60/146,222  
PRIOR FILING DATE: 1999-07-28  
PRIOR APPLICATION NUMBER: PCT/US99/20594  
PRIOR FILING DATE: 1999-09-08  
PRIOR APPLICATION NUMBER: PCT/US99/20944  
PRIOR FILING DATE: 1999-09-13  
PRIOR APPLICATION NUMBER: PCT/US99/21090  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/21547  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/23089  
PRIOR FILING DATE: 1999-10-05



; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 261
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-09-907-613-261

Query Match
Best Local Similarity 7.4%; Score 70.5; DB 9; Length 383;
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;

QY 7 GPVIGKSLTLKLPAYWELSGFHQLTDQWAIHYSYKYTEWSRKEKLRGKY-----56
Db 66 GPOCHKGT---PLPTVEEAKQY-----LSYETLYANGSRRTETQGVILSSGGDAQ 114

QY 57 --QDGSYEAFTKKEEYKNSRFAT-GTYSLSN-----DALTLRAGLAYDKAASKTHLSAS 109
Db 115 HRDSSGSKRRKRIQYGVDSRESIFGKDFLLNYPSTSVKLSGTGCTGLVAEKHVLTA 174

QY 110 --IPDTRMWSYIGATY-KETPNLSV-----DVGFAHLRGKK 143
Db 175 HCITHD-----GKTVVKGTKQLRVGFLKPKFKDGGRGANDSTSAMPEOMKQFQWIRVKR 226

QY 144 KHEVETQNIKGLLLEADYTTKATANLYGLNLNY 177
Db 227 THVP-----KGWI-----KGNANDIGMDYDY 247

RESULT 34
US-09-907-942-261
; Sequence 261, Application US/09907942
; Publication No. US20030027146A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey

; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,942
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 261
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-09-907-942-261

Query Match
Best Local Similarity 7.4%; Score 70.5; DB 9; Length 383;
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;

QY 7 GPVIGKSLTLKLPAYWELSGFHQLTDQWAIHYSYKYTEWSRKEKLRGKY-----56
Db 66 GPOCHKGT---PLPTVEEAKQY-----LSYETLYANGSRRTETQGVILSSGGDAQ 114

QY 57 --QDGSYEAFTKKEEYKNSRFAT-GTYSLSN-----DALTLRAGLAYDKAASKTHLSAS 109
Db 115 HRDSSGSKRRKRIQYGVDSRESIFGKDFLLNYPSTSVKLSGTGCTGLVAEKHVLTA 174

QY 110 --IPDTRMWSYIGATY-KETPNLSV-----DVGFAHLRGKK 143
Db 175 HCITHD-----GKTVVKGTKQLRVGFLKPKFKDGGRGANDSTSAMPEOMKQFQWIRVKR 226

QY 144 KHEVETQNIKGLLLEADYTTKATANLYGLNLNY 177
Db 227 THVP-----KGWI-----KGNANDIGMDYDY 247

RESULT 35
US-10-173-700-50
; Sequence 50, Application US/10173700
; Publication No. US20030027262A1
; GENERAL INFORMATION:

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C14
; CURRENT APPLICATION NUMBER: US/10/173,700
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 50
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-700-50

Query Match      7.4%; Score 70.5; DB 9; Length 383;
Best Local Similarity 23.8%; Pred. No. 36;
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;

QY 7 GPYIGKSLTLKLPAYWELSGFHQLTQDWAHYSKYKTEWSRFKELRGKY----- 56
DB 66 GPQCHKGT---PLPYEEAKQY-----LSYETLYANGSRTEYQVGIYILSSSGDGAQ 114
QY 57 --QDGSQGEAFTKKEEYKDNSRFAT-GTYSLSN-----DALTLRAGLAYDKAASKTHLSAS 109
DB 115 HRDSSGSKSRKRQIYGYDSRFSIFGKDFLLNYPFSTSVKLSCTGTGLVAEKHVLTA 174
QY 110 --IPDTRMWSIGATY-KFTPNLSV-----DVGFAHLRGKK 143
DB 175 HCIHD-----GKTYVKGTKLRVGLPKPKFDGGRGANDSTSAMPEQMKFQWIRVKR 226
QY 144 KHVETQNIKGLLLVEADYTTKATANLYGLNLNY 177
DB 227 THVP-----KGWI-----KGNANDIGMDYDY 247

RESULT 36
US-10-174-572-50
; Sequence 50, Application US/10174572
; Publication No. US20030027263A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C31
; CURRENT APPLICATION NUMBER: US/10/174,579
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 50
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-572-50

Query Match      7.4%; Score 70.5; DB 9; Length 383;
Best Local Similarity 23.8%; Pred. No. 36;
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;

QY 7 GPYIGKSLTLKLPAYWELSGFHQLTQDWAHYSKYKTEWSRFKELRGKY----- 56
DB 66 GPQCHKGT---PLPYEEAKQY-----LSYETLYANGSRTEYQVGIYILSSSGDGAQ 114
QY 57 --QDGSQGEAFTKKEEYKDNSRFAT-GTYSLSN-----DALTLRAGLAYDKAASKTHLSAS 109
DB 115 HRDSSGSKSRKRQIYGYDSRFSIFGKDFLLNYPFSTSVKLSCTGTGLVAEKHVLTA 174
QY 110 --IPDTRMWSIGATY-KFTPNLSV-----DVGFAHLRGKK 143
DB 175 HCIHD-----GKTYVKGTKLRVGLPKPKFDGGRGANDSTSAMPEQMKFQWIRVKR 226
QY 144 KHVETQNIKGLLLVEADYTTKATANLYGLNLNY 177
DB 227 THVP-----KGWI-----KGNANDIGMDYDY 247

RESULT 37
US-10-174-579-50
; Sequence 50, Application US/10174579
; Publication No. US20030027264A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C31
; CURRENT APPLICATION NUMBER: US/10/174,579
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 50
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-579-50

Query Match      7.4%; Score 70.5; DB 9; Length 383;
Best Local Similarity 23.8%; Pred. No. 36;
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;

QY 7 GPYIGKSLTLKLPAYWELSGFHQLTQDWAHYSKYKTEWSRFKELRGKY----- 56
DB 66 GPQCHKGT---PLPYEEAKQY-----LSYETLYANGSRTEYQVGIYILSSSGDGAQ 114
QY 57 --QDGSQGEAFTKKEEYKDNSRFAT-GTYSLSN-----DALTLRAGLAYDKAASKTHLSAS 109
DB 115 HRDSSGSKSRKRQIYGYDSRFSIFGKDFLLNYPFSTSVKLSCTGTGLVAEKHVLTA 174
QY 110 --IPDTRMWSIGATY-KFTPNLSV-----DVGFAHLRGKK 143
DB 175 HCIHD-----GKTYVKGTKLRVGLPKPKFDGGRGANDSTSAMPEQMKFQWIRVKR 226
QY 144 KHVETQNIKGLLLVEADYTTKATANLYGLNLNY 177
DB 227 THVP-----KGWI-----KGNANDIGMDYDY 247

RESULT 38
US-10-174-572-50
; Sequence 50, Application US/10174572
; Publication No. US20030027263A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C40
; CURRENT APPLICATION NUMBER: US/10/174,572
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 50
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-572-50
```

```
US-10-174-582-50
; Sequence 50, Application US/10174582
; Publication No. US20030027265A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C36
; CURRENT APPLICATION NUMBER: US/10/174,582
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 50
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-582-50

Query Match      7.4%; Score 70.5; DB 9; Length 383;
Best Local Similarity 23.8%; Pred. No. 36;
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;

QY 7 GPYIGKSLTLKLPAYWELSGFHOLTQDWAHYSKYKTEWSRFRKELRGKY-----56
DB 66 GPQCHKGT---PLPTYEAKQY-----LSYETLVANGSRSTETQVGIYILSSSGDGAQ 114
QY 57 --QDGSYEAFTKKEEYKDNSRFAT-GTYSLSN----DALTLRAGLAYDKAAKSTHLSAS 109
DB 115 HRDSSGSKSRKQIYGYDSRFSIFGKDFLLNYPFSTSVKLSGTCTGLVAEKHVLTA 174
QY 110 --IPDTRMYSIGATY-KETPNLSV-----DVGFAHLRGKK 143
DB 175 HCIHD-----GKTVYKGTQKLRVGLPKPKDGGRGANDSTSAMPEQMKFQWIRVKR 226
QY 144 KHVETONIKGLLVEADYTTKATANLYGLNLNY 177
DB 227 THVP-----KGWI-----KGNANDIGMDYDY 247

RESULT 39
US-10-174-588-50
; Sequence 50, Application US/10174588
; Publication No. US20030027266A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C28
; CURRENT APPLICATION NUMBER: US/10/174,588
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 50
```

```
US-10-175-739-50
; Sequence 50, Application US/10175739
; Publication No. US20030027267A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C46
; CURRENT APPLICATION NUMBER: US/10/175,739
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 50
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-739-50

Query Match      7.4%; Score 70.5; DB 9; Length 383;
Best Local Similarity 23.8%; Pred. No. 36;
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;

QY 7 GPYIGKSLTLKLPAYWELSGFHOLTQDWAHYSKYKTEWSRFRKELRGKY-----56
DB 66 GPQCHKGT---PLPTYEAKQY-----LSYETLVANGSRSTETQVGIYILSSSGDGAQ 114
QY 57 --QDGSYEAFTKKEEYKDNSRFAT-GTYSLSN----DALTLRAGLAYDKAAKSTHLSAS 109
DB 115 HRDSSGSKSRKQIYGYDSRFSIFGKDFLLNYPFSTSVKLSGTCTGLVAEKHVLTA 174
QY 110 --IPDTRMYSIGATY-KETPNLSV-----DVGFAHLRGKK 143
DB 175 HCIHD-----GKTVYKGTQKLRVGLPKPKDGGRGANDSTSAMPEQMKFQWIRVKR 226
QY 144 KHVETONIKGLLVEADYTTKATANLYGLNLNY 177
DB 227 THVP-----KGWI-----KGNANDIGMDYDY 247

RESULT 40
US-10-175-739-50
; Sequence 50, Application US/10175739
; Publication No. US20030027267A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C46
; CURRENT APPLICATION NUMBER: US/10/175,739
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 50
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-739-50

Query Match      7.4%; Score 70.5; DB 9; Length 383;
Best Local Similarity 23.8%; Pred. No. 36;
Matches 51; Conservative 22; Mismatches 66; Indels 75; Gaps 11;

QY 7 GPYIGKSLTLKLPAYWELSGFHOLTQDWAHYSKYKTEWSRFRKELRGKY-----56
DB 66 GPQCHKGT---PLPTYEAKQY-----LSYETLVANGSRSTETQVGIYILSSSGDGAQ 114
QY 57 --QDGSYEAFTKKEEYKDNSRFAT-GTYSLSN----DALTLRAGLAYDKAAKSTHLSAS 109
DB 115 HRDSSGSKSRKQIYGYDSRFSIFGKDFLLNYPFSTSVKLSGTCTGLVAEKHVLTA 174
QY 110 --IPDTRMYSIGATY-KETPNLSV-----DVGFAHLRGKK 143
DB 175 HCIHD-----GKTVYKGTQKLRVGLPKPKDGGRGANDSTSAMPEQMKFQWIRVKR 226
QY 144 KHVETONIKGLLVEADYTTKATANLYGLNLNY 177
DB 227 THVP-----KGWI-----KGNANDIGMDYDY 247
```

Db 227 THVP-----KGWI-----KGNANDIGMDYDY 247

Search completed: May 12, 2003, 09:46:36  
Job time : 25 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: May 12, 2003, 09:39:14 ; Search time 17 Seconds  
(without alignments)  
1012.239 Million cell updates/sec

Title: US-09-995-493-52  
Perfect score: 957  
Sequence: 1 QHNGVLGPYIGKSLTKLP.....ADYTTKATANLYGLNLNRYF 179  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:.\*  
1: p1r1.\*  
2: p1r2.\*  
3: p1r3.\*  
4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	570.5	59.6	450	2 B40183	outer membrane pro
2	559	58.4	455	2 A40183	outer membrane pro
3	549	57.4	459	2 A28787	outer membrane pro
4	549	57.4	459	2 F84065	outer membrane pro
5	421	44.0	412	2 G82406	long-chain fatty a
6	372.5	38.9	423	2 A80334	probable long-chain
7	338	35.3	437	2 A80805	long-chain fatty a
8	334	34.9	432	2 B82250	long-chain fatty a
9	332	34.7	448	2 F65007	long-chain fatty a
10	332	34.7	448	2 C31032	long-chain fatty a
11	332	34.7	448	2 D85876	long-chain fatty a
12	267.5	28.0	428	2 B82248	long-chain fatty a
13	207	21.6	447	2 E82729	outer membrane pro
14	194.5	20.3	432	2 E87508	outer membrane pro
15	190.5	19.9	463	2 A83073	probable outer mem
16	167	17.5	424	2 H83484	probable outer mem
17	144.5	15.1	466	2 A81239	outer membrane pro
18	143.5	15.0	532	2 G83424	hypothetical prote
19	142.5	14.9	464	2 H82011	probable outer mem
20	132.5	13.8	403	2 C70385	hypothetical prote
21	111	11.6	308	2 H70355	hypothetical prote
22	108	11.3	587	2 G84624	outer membrane pro
23	106.5	11.1	453	2 S33995	Tox protein - Pse
24	102	10.7	587	2 C71889	probable outer mem
25	100	10.4	753	2 D83081	probable outer mem
26	96.5	10.1	708	2 B81038	TonB-dependent rec
27	94	9.8	572	2 A81209	N-acetylmuramoyl-L
28	92.5	9.7	2425	2 D69426	surface layer prot
29	91.5	9.6	730	2 B83592	hypothetical prote

30 90.5 9.5 362 2 D90755 outer membrane pro  
31 90.5 9.5 362 2 B85619 outer membrane pro  
32 90.5 9.5 875 2 S49608 outer membrane u  
33 88.5 9.2 522 2 G81243 peptide methionine  
34 88 9.2 479 2 A71273 hypothetical prote  
35 87 9.1 182 2 B98110 probable virulence  
36 85.5 8.9 177 2 E85955 outer membrane por  
37 85.5 8.9 362 1 MMECF vitamin B12 recept  
38 85.5 8.9 611 2 A82358 penicillin-binding  
39 85.5 8.9 916 2 G97053 hypothetical prote  
40 85 8.9 221 2 F84148 adhesion and penet  
41 84.5 8.8 1457 2 D81019 hemeoglobin-haptog  
42 84 8.8 810 2 A81965 choline binding pr  
43 83.5 8.7 332 2 H95043 ADP-ribosylarginin  
44 83.5 8.7 362 2 A38135 hypothetical prote  
45 83 8.7 440 2 D71715

ALIGNMENTS

RESULT 1

B40183

outer membrane protein P1 precursor, subtype 6U - Haemophilus influenzae (strain 8358)  
C:Species: Haemophilus influenzae  
C:Date: 28-Oct-1992 #sequence\_revision 30-Jan-1993 #text\_change 12-Dec-1997  
C:Accession: B40183  
R:Musson Jr., R.; Grass, S.; Einhorn, M.; Bailey, C.; Newell, C.  
Infect. Immun. 57, 3300-3305, 1989.  
A:Title: Comparative analysis of the structures of the outer membrane protein P1 gene  
A:Reference number: A40183; MUID:90035394; PMID:2572549  
A:Accession: B40183  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-450 <MUN>  
A:Cross-references: GB:M27683  
C:Superfamily: long-chain fatty acid transport protein fadL  
C:Keywords: membrane protein  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-450/Product: outer membrane protein P1 #status predicted <MAT>

Query Match 59.6%; Score 570.5; DB 2; Length 450;

Best Local Similarity 62.7%; Pred. No. 4.8e-45;

Matches 106; Conservative 30; Mismatches 30; Indels 3; Gaps 2;

QY 12 KGSLLTLKPAYWELSGFHQLTQWATHYKYKTEHSRFEKELRGKYODSGYEFTKKEEY 71

DB 284 KGNLTLLKLPDYLELSGFHQLTDKFAVHYKYTHWSRLTKLHASFEDGK--KAFDKELQY 341

QY 72 KONSRAFGITTYSLNDALTLRAGLAYDKAASKTHLSASIPDTRMWSYSGATYKFTPNLS 131

DB 342 SNNRSIALGASYNLDEKLTLAGIADYDQAASRHHRSAAIPDTRWTYSLGATYKFTPNLS 401

QY 132 VDVGFAHLRGKKHFVETONIK--GLLLVEADYTTKATANLYGLNLNRYF 179

DB 402 VDLGAYLKGKVKHFEKAQQAAGGFTTTTANTYTSQAHANLYGLNLNRYF 450

RESULT 2

A40183

outer membrane protein P1 precursor, subtype 3L - Haemophilus influenzae (strain 1613)  
C:Species: Haemophilus influenzae  
C:Date: 28-Oct-1992 #sequence\_revision 30-Jan-1993 #text\_change 20-Aug-1999  
C:Accession: A40183  
R:Musson Jr., R.; Grass, S.; Einhorn, M.; Bailey, C.; Newell, C.  
Infect. Immun. 57, 3300-3305, 1989.  
A:Title: Comparative analysis of the structures of the outer membrane protein P1 gene  
A:Reference number: A40183; MUID:90035394; PMID:2572549  
A:Accession: A40183  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-455 <MUN>  
A:Cross-references: GB:M63151; GB:M27682; NID:g148954; PIDN:AAA24991.1; PID:g148955



ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AH0334  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-423 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC92983.1; PID:gi5980722; GSPDB:GN00175  
C:Genetics:  
A:Gene: fadL  
C:Superfamily: long-chain fatty acid transport protein fadL

Query Match 38.9%; Score 372.5; DB 2; Length 423;  
Best Local Similarity 44.6%; Pred. No. 9.4e-27;  
Matches 79; Conservative 23; Mismatches 70; Indels 5; Gaps 3;

QY 4 GVLGPYIGKSLTLKPAYWELSGPHQLTDQWAIHYSYKYTEWSRFKELRGYDQSGGYE 63  
DB 251 GGLGGKTVPGLTLNLPAYWEVSGYNKVAPOWAIHYSMAITWSSFKELKATASNGD--V 308  
QY 64 AFTKKEEKDNRFAIGTYSLNDALTLRAGLAYDKAASKT-HLSASIPDTRMYSGICA 122  
DB 309 LFDKHEGFADVRIALGTTYYDDNNWTFRTGIAFDSPIPAGNRSISIPDQRFWSAGT 368  
QY 123 TYKFTPNLSVDVGFALHKGKKHFEVETQNIKGLLLVEADYTTKATANLYGLNLNRYF 179  
DB 369 TVAFKNASVDVGIAMKQGNYSITEKTPAPSNITVE--FNSKGSAMLYGVNFNYTF 423

RESULT 7  
AB0805  
long-chain fatty acid transport protein precursor STY2623 [imported] - Salmonella enteri  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AE0805  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov A:Reference number: AB0502; PMID:11677608  
A:Accession: AE0805  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-437 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD07623.1; PID:gl6503614; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY2623  
C:Superfamily: long-chain fatty acid transport protein fadL

Query Match 35.3%; Score 338; DB 2; Length 437;  
Best Local Similarity 42.9%; Pred. No. 1.5e-23;  
Matches 73; Conservative 26; Mismatches 55; Indels 16; Gaps 5;

QY 13 GSLTLKPAYWELSGPHQLTDQWAIHYSYKYTEWSRFKELRGYDQSGGYEAFKKKEEK 72  
DB 281 GYLTNLNPEWVSGYNRRVAPOWAIHYSLAYTWSQFQELKAK--STAGDTLEKHEGFR 338  
QY 73 DNSRFAIGTYSLNDALTLRAGLAYDKA-ASKTHLSASIPDTRMYSGICATYKFTPNLS 131  
DB 339 DAYRIALGTTYYDDNNWTFRTGIAFDSPVPAQNRSISIPDQRFWSAGTTFYAFNKDAS 398  
QY 132 VDVGFAHLRGKKHFEVETQNIKGLLLVEADY--TTKATANLYGLNLNRYF 179  
DB 399 VDVGVSVMHG-----QSVK---INEGPQFESECKAWLFGTGNFYAF 437

RESULT 8  
B82250  
long-chain fatty acid transport protein VC1042 [imported] - Vibrio cholerae (strain N169

C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: B82250  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: AB2035; MUID:20406833; PMID:10952301  
A:Accession: B82250  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-432 <HEI>  
A:Cross-references: GB:AE004185; GB:AE003852; NID:g9655497; PIDN:AAF94201.1; GSPDB:GN A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC1042  
A:Map position: 1  
C:Superfamily: long-chain fatty acid transport protein fadL

Query Match 34.9%; Score 334; DB 2; Length 432;  
Best Local Similarity 41.3%; Pred. No. 3.6e-23;  
Matches 71; Conservative 30; Mismatches 63; Indels 8; Gaps 4;

QY 13 GSLTLKPAYWELSGPHQLTDQWAIHYSYKYTEWSRFKELRGYDQSGGYEAFKKKEEK 72  
DB 264 GMSVTLPTAELASPHQLNDQWAIHASINWTDWSSFKELTAVFEKSD--LIKSENWE 320  
QY 73 DNSRFAIGTYSLNDALTLRAGLAYD-KAASKTHLSASIPDTRMYSGICATYKFTPNLS 131  
DB 321 DNYRFALGTTYYQYDAKLALRAGVAYDTSVDDKNRTATIPETDRTWVSGSVVATPQLT 380  
QY 132 VDVGFAHLRGKKHFEVETQNI--KGLLLVEADYTTKATAN--LYGLNLNRYF 179  
DB 381 LDAGFTYIIFAKDATINEPRDASQDTAAATCGAFTGNVSGNVLIGVQANRYF 432

RESULT 9  
F65007  
long-chain fatty acid transport protein fadL precursor - Escherichia coli (strain K-1  
N:Alternate names: FLP protein  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: F65007; A39126; S02829  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M., A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: F65007  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-448 <BLAT>  
A:Cross-references: GB:AE000323; GB:U00096; NID:gl788684; PIDN:AAC75404.1; PID:gl7886 A:Experimental source: strain K-12, substrain MG1655  
R:Black, P.N.  
J. Bacteriol. 173, 435-442, 1991  
A:Title: Primary sequence of the Escherichia coli fadL gene encoding an outer membran A:Reference number: A39126; MUID:91100327; PMID:1987139  
A:Accession: A39126  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-177, 'V', 179-287, 'R', 289-448 <BLA>  
A:Cross-references: GB:M60607; GB:M37714; NID:gl145909; PIDN:AAA64433.1; PID:gl145910 R:Said, B.; Ghosh, C.R.; Vu, L.; Nunn, W.D.  
Mol. Microbiol. 2, 363-370, 1988  
A:Title: Nucleotide sequencing and expression of the fadL gene involved in long-chain A:Reference number: S02829; MUID:88288050; PMID:2840553  
A:Accession: S02829  
A:Molecule type: DNA  
A:Residues: 66-177, 'R', 179, 'P', 181-287, 'R', 289-448 <SAI>  
A:Cross-references: EMBL:Y00552; NID:g41371; PIDN:CAA68630.1; PID:g41372  
A:Note: the authors translated the codon CTG for residue 162 as Glu









A:Accession: C70385  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-403 <AQ>  
A:Cross-References: GB:AE000717; NID:g2983492; PIDN:AAC07076.1; PID:g2983498; GB:AE00065  
A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: aq\_985

Query Match 13.8%; Score 132.5; DB 2; Length 403;  
Best Local Similarity 25.4%; Pred. No. 0.00015;  
Matches 44; Conservative 39; Mismatches 73; Indels 17; Gaps 8;  
QY 13 GSUTLKLPAWELSGFHQTDQWAIHYSYKTEWRSFKELRGKYDQSGYEAFK--KEE 70  
DB 242 GDVSVPLPAELRISGSY--WNNTTTELRTFWSSDYDLDFNYKDPVAENQFGKPIKD 300  
QY 71 YKNSRFAIGTYSUNDALTLAGLAYDKA--ASKTHLSASIPDTRMW-YSIGATYKFT 127  
DB 301 WEDVNTYRFLQHKFGDSFKGLGIAYSESPISKS-LGFELPEKSVWILSLGGIYNLT 359  
QY 128 PNLSDVDVGFALH-RCKKKHFVETONIKGLLLVEADYTTKATANLYCLNLYRF 179  
DB 360 KMW--EVGFSYLLAKDSRKVDNQRIKGEF-----SDISAHFLTSLVGAKF 403

RESULT 21  
H70355  
hypothetical protein aq\_627 - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 18-Aug-2000  
C:Accession: H70355  
V:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov

Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: H70355  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-308 <AQ>  
A:Cross-References: GB:AE000659; NID:g2983238; PIDN:AAC06840.1; PID:g2983244; GB:AE00065  
A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: aq\_627  
C:Superfamily: Aquifex aeolicus hypothetical protein aq\_627

Query Match 11.6%; Score 111; DB 2; Length 308;  
Best Local Similarity 22.4%; Pred. No. 0.01;  
Matches 37; Conservative 25; Mismatches 55; Indels 48; Gaps 7;  
QY 42 KYTEWSRFRKRLCKYDQSGYEAFKKEEYKNSRFAIGTYSUNDALTLAGLAYDKA- 100  
DB 165 RTIDWS-----GADGYGDF-----DWEDQWVFLGVYKPMDSLALRAGWNYGKTP 210  
QY 101 -ASKTHLSASIP-----DTRMWY-----SIGATYKFTPNLSVDV 134  
DB 211 IRDKNDLPMTKNSVPSEDTFESDQLYFNVLGVPFAEHHLTGLNWKITPNENIDL 270  
QY 135 GFALHKGKKHFFVETONIKGLLLVEADYTTKATANLYCLNLYRF 179  
DB 271 SYVVALPKT---VTAKAAG-----AGEVSTKMVQHSIGVGLNNAF 308

RESULT 22  
G64624  
outer membrane protein P1 - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori  
C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
A:Accession: G64624  
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne  
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.

Nature 388, 539-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, A.;  
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A:Reference number: A64520; MUID:97394467; PMID:9252185  
A:Accession: G64624  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-587 <TOM>  
A:Cross-References: GB:AE000595; GB:AE000511; NID:g2313969; PIDN:AAD07886.1; PID:g231

Query Match 11.3%; Score 108; DB 2; Length 587;  
Best Local Similarity 23.3%; Pred. No. 0.044;  
Matches 49; Conservative 29; Mismatches 78; Indels 54; Gaps 10;  
QY 6 LGPYTG---KGSLLT--KLPAWELSGFHQTDQWAIHYSYKTEWRSFKEL----- 52  
DB 396 LGPSGVSUTKSLNINVSLOTLSLAYAHQFFKQHLRIEGVFERTFWSQGNKFLVTPDP 455  
QY 53 -RCKYDQSGYEAFKKEE-----YKNSRFAIGTYSUNDALTLAGLAYDKA--ASKTHLSASIPDTRMW-YSIGATYKFT 92  
DB 456 ANATYKGLSCTVASLDSETLKKMVLNFKSVNMGAGWRDNTTFRLGVTY-MGKSLRLM 514  
QY 93 AGLAYDKAASKTHLSASIPDTRMWYSIGATYKFTPNLSVDVGF-----HLRCKKKHFFVET 149  
DB 515 GAILYDQAPSQD-AIGIPDSNGYTVAFGTYKYNFR---GFDLGVAGSFTFKNSRSLYQS 570  
QY 150 QNIKGLLLVEADYTTKATANLYCLNLYRF 179  
DB 571 PNIGQL-----RIFSASLGYRW 587

RESULT 23  
S53995  
ToxX protein - Pseudomonas putida  
C:Species: Pseudomonas putida  
C:Date: 15-Jul-1995 #sequence\_revision 17-Nov-1995 #text\_change 08-Oct-1999  
C:Accession: S53995  
R:Wang, Y.; Rawlings, M.; Gibson, D.T.; Labbe, D.; Bergeron, H.; Brousseau, R.; Lau,  
Mol. Gen. Genet. 246, 570-579, 1995  
A:Title: Identification of a membrane protein and a truncated Lysr-type regulator ass  
A:Reference number: S53993; MUID:95214617; PMID:7535376  
A:Accession: S53995  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-453 <WAN>  
A:Cross-References: GB:U18304; NID:g747633; PIDN:AAC43318.1; PID:g747636

Query Match 11.1%; Score 106.5; DB 2; Length 453;  
Best Local Similarity 27.1%; Pred. No. 0.044;  
Matches 38; Conservative 18; Mismatches 59; Indels 25; Gaps 6;  
QY 13 GSUTLK---LPAYWELSGFHQTDQWAIHYSYKTEWRSFKELRGKYDQSGYEAFKTK- 68  
DB 288 GDIRVAKFMFASLTTLGLAHQFNFRWVVAADIKRAYW-----GDVMDSMNV-AFISQL 339  
QY 69 -----BEYKNSRFAIGTYSUNDALTLAGLAYDKAASKTHLSAS-IPOTDRMWYS 119  
DB 340 GGIDVALPHRYQDITVASIGTAYKYNNDLTRAGYSYAQAQLDSELILPILPAYLKRHVT 399  
QY 120 IGATYKFTP-----NLSVDVVG 135  
DB 400 FGGEYDFDKDSRLNLAISFG 419

RESULT 24  
C71889  
probable outer membrane protein - Helicobacter pylori (strain J99)  
C:Species: Helicobacter pylori  
A:Variety: strain J99  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 08-Oct-1999  
A:Accession: C71889  
R:Alm, R.A.; Ling, L.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.  
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F

Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
A:Reference number: A71800; MUID:99120557; PMID:9923682  
A:Accession: C71889  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-587 <ARN>  
A:Cross-references: GB:AE001508; GB:AE001439; MID:g4155338; PIDN:AA06357.1; PID:g415534  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: jhp0777

Query Match 10.7%; Score 102; DB 2; Length 587;  
Best Local Similarity 25.5%; Pred. No. 0.16;  
Matches 42; Conservative 23; Mismatches 62; Indels 38; Gaps 8;

QY 6 LGPYIG----KGSITL--KLPAWELSGFHQ-LTDQWAIHYSKYTWSRFKEL-----52  
Db 396 LGPSLGSVLTKGSUNINVSLPQTLALAYAHOFKDLRLRVGVFTWFOGKFLVTPDF 455  
QY 53 -RGKYQDGSVEAFTKKEE-----YKDNRSFAIGTYSINDALTLR 92  
Db 456 ANATYKGLSCTVSLDSETLKKVGLANFKSVNMAGWRDNTFRGLGTY-MCKSLRLM 514  
QY 93 AGLAYKAAASKTHLSASIPDTRMWSIGATYKFTPNLSVDVGF 137  
Db 515 GAIDYDQAPSPQD-AIGIPDSNGYTVAFGTYKYNFR---GFDLGA 555

RESULT 25  
D83081  
Probable outer membrane receptor for iron transport PA4514 [imported] - Pseudomonas aeru  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 14-Sep-2001  
C:Accession: D83081  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: D83081  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-753 <STO>  
A:Cross-references: GB:AE004865; GB:AE004091; MID:g950752; PIDN:AA07902.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA4514  
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol

Query Match 10.4%; Score 100; DB 2; Length 753;  
Best Local Similarity 27.5%; Pred. No. 0.33;  
Matches 33; Conservative 20; Mismatches 53; Indels 14; Gaps 6;

QY 27 GFH-QLTDQWAIHYSKYTWSRFKELRGKYQDGSVEATKKEEYKDNRSFAIGTYSYL 85  
Db 618 GFNGKLETKWKFVGGYTYLD-SEIKSTVKSDEGN-----KMPQTAQNNFTLWTYDL 669  
QY 86 NDALTTRAGLAY-DKAASKTHLSASIPDTRMW-YSIGATYKFTPNLSVDVGFALRGKK 143  
Db 670 LQNFITGGTYYDKVQGNANTSTYIPS---YWRDAMASYKSKVNDLQINQNLTKR 726

RESULT 26  
B81038  
TonB-dependent receptor NMB1829 [imported] - Neisseria meningitidis (strain MC58 serogr  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: B81038  
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.

Science 287, 1809-1815, 2000.  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;  
A:Title: Complete genome sequence of Neisseria meningitidis sdrgroup B strain MC58.  
A:Reference number: A81000; MUID:20175755; PMID:10710307  
A:Accession: B81038  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-708 <TET>  
A:Cross-references: GB:AE002532; GB:AE002098; MID:g7227078; PIDN:AAF42164.1; PID:g7222  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB1829

Query Match 10.1%; Score 96.5; DB 2; Length 708;  
Best Local Similarity 24.4%; Pred. No. 0.65;  
Matches 44; Conservative 25; Mismatches 70; Indels 41; Gaps 9;

QY 6 LGPYIGKGSITLKLPAWELSGFHQDQWAIHYSKYTWSRFKELRGKYQDGSVEAF 65  
Db 564 LGKRVMEGVET-----EISG--AMTPKQIHAGYSYLH-SQIKTASNRDEG----IF 609  
QY 66 TKREYKDNRSFAIGTYSINDALTTRAGLAYDKAASKTHLSASIPDTRMWSIGATYK 125  
Db 610 LMPKHSAN----LWTTYQVTSGLTIGGV---NMSGITSSAGIHAGGYATPDAMAYR 662  
QY 126 FTPNLSVDVGFALRGKKHFEVETQNIKGLLVEADYTTKATANLYG-----LNLNRYF 179  
Db 663 FTPKCLKLIQINADNI--PNRHY-----ARVSESTFNPICERSLNTANLAYSF 708

RESULT 27  
AD1209  
N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) lmo1076 - Listeria monocytogenes (st  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 31-Dec-2001  
C:Accession: AD1209  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl,  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maitournam, A.;  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
A:Title: Comparative genomics of Listeria species.  
A:Reference number: A81077; MUID:21537279; PMID:11679669  
A:Accession: AD1209  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-572 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAC99154.1; PID:gl6410478; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo1076  
C:Keywords: hydrolase

Query Match 9.8%; Score 94; DB 2; Length 572;  
Best Local Similarity 26.1%; Pred. No. 0.84;  
Matches 57; Conservative 30; Mismatches 67; Indels 64; Gaps 12;

QY 4 GVLGPYIGKGSITLKLPAWELSGFHQDQWAIHYSKYTWSRFKELRGKYQDGSY 62  
Db 140 GIKGSYNGK-SVTMKTWESDSKGWQINANFAKYPSPHESKESLEDNAKLRNGPSWDSYY 198;  
QY 63 EAFYTK--EYKD-----NSRFAIGTYS--LN-----D 87  
Db 199 KGAWRENATKYKDATALMGQRYATDNTYASKNLTLISSNLTQYDLYTIKQKNVSED 258  
QY 88 ALTLRA-----GLAYDKAASKTHLSASIPDTR-----MWYSIGATYKFTPNLSVD 133  
Db 259 AKVVKADGHGVSGIYNTWSAASAKKJSTGAPYNNKVKILKEGTTSRGTVWVQSLNNKV- 317  
QY 134 VGFAHLRGKKHFFV---ETQNIKGLLVEADYTTKATA 168  
Db 318 IGWM-----DKRAFVYYPKATNVKTLNL-----TGKITA 346

RESULT 28  
D69426  
surface layer protein B (slgB-2) homolog - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 21-Jul-2000  
C:Accession: D69426  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
R: Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 350, 364-370, 1997  
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
A:Reference number: A69250; MUID:98049343; PMID:9389475  
A:Accession: D69426  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2425 <KLE>  
A:Cross-references: GB:AE001006; GB:AE000782; NID:g2689329; PIDN:AAB89834.1; PID:g264915

Query Match 9.7%; Score 92.5; DB 2; Length 2425;  
Best Local Similarity 25.2%; Pred. No. 7.3;  
Matches 36; Conservative 17; Mismatches 43; Indels 47; Gaps 8;

QY 3 NGVLGPGYIGKSLTKLKPAYW-----ELSGFHQLTDQWAIHVSRYKTEWSRFKELRGKY 56  
DB 998 NIIGPGYLGNN-----YHDIYAGEDVAGDGLGDTLLPNS-----GGNI 1037

QY 57 QGSGYEAFTK-KEEYKNSRFAT-----GTTYSLNDALTFRAGLAYDKAASKTHLSASIP 111  
DB 1038 QNGGDYHPLTNPNDFRAPAIYVSPVEGRYTSAN-----YVLEVYSPD-----P 1083

QY 112 DTRMWSYI--GATYKFTPNLSV 132  
DB 1084 DVDRMWSLNSGANVTFPTNTTI 1106

RESULT 29  
B83592  
hypothetical protein PA0434 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: B83592  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: B83592  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-730 <STO>  
A:Cross-references: GB:AE004480; GB:AE004091; NID:g9946284; PIDN:AAG03823.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0434

Query Match 9.6%; Score 91.5; DB 2; Length 730;  
Best Local Similarity 25.4%; Pred. No. 2;  
Matches 34; Conservative 16; Mismatches 45; Indels 39; Gaps 7;

QY 31 LTDQWAIHVSRYKTEWSRFKELR-----GKYDGSYEAFTKKE-----EYKD 73  
DB 377 LASDWALGFDYSINKQTRFPSTASPGFTVDPASPEPGHFYDLPGRPCGHRKDRSNEVRT 436

QY 74 NSRFAIGTYSLNDALTFRAGLAYDKAASKTHLSA-----IPDPTD-----RMWYSI-- 120  
DB 437 SALFA-ENRGLGTDLSLVLTGLRYD-----HLDLDRNHRVPDKNPAHFERRDVTG 489

QY 121 --GATYKFTPNLSV 132

DB 490 RAGLVYQFTPHANV 503  
| :|:|:|:|  
RESULT 30  
D90755  
outer membrane protein la Ecs1012 [imported] - Escherichia coli (strain O157:H7, subs  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: D90755  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g  
A:Reference number: A99629; MUID:21156231; PMID:11250796  
A:Accession: D90755  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-362 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA034435.1; PID:g13360471; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: Ecs1012  
C:Superfamily: outer membrane protein phoE

Query Match 9.5%; Score 90.5; DB 2; Length 362;  
Best Local Similarity 25.9%; Pred. No. 1;  
Matches 37; Conservative 22; Mismatches 59; Indels 25; Gaps 6;

QY 4 GVLGPGYIGKSLTKLKPAYWELSGFHQLTDQW-----IHVSRYKTEWSRFKELRG 54  
DB 208 GIVGAY-GAADRNLQEA--QLLNGKKAQWATGLKYDANNIYLAANYGETRNATPIIN 264

QY 55 KYDGSYEAFTKKEEYKNSRFATIGTYSLNDALTFRAGLAYDKAASKTHLSASIPDPTD 114  
DB 265 KFTNISGFEANKQDVLVLAQYQDFG-----LRPSIATYKSKAKD--VEGIGDVD 312

QY 115 RM-WYSIGATYKFTPNLSVDVGF 136  
DB 313 LVNYFEVGATYFENKMWSTYVDY 335

RESULT 31  
B85619  
outer membrane protein la (Ia,b,F) [imported] - Escherichia coli (strain O157:H7, sub  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: B85619  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasne4, J.D.; Rose, D.J.; May  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: B85619  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-362 <STO>  
A:Cross-references: GB:AE005174; NID:g12514104; PIDN:AAG55414.1; GSPDB:GN00145; UNCP:  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: ompF  
C:Superfamily: outer membrane protein phoE

Query Match 9.5%; Score 90.5; DB 2; Length 362;  
Best Local Similarity 25.9%; Pred. No. 1;  
Matches 37; Conservative 22; Mismatches 59; Indels 25; Gaps 6;

QY 4 GVLGPGYIGKSLTKLKPAYWELSGFHQLTDQW-----IHVSRYKTEWSRFKELRG 54  
DB 208 GIVGAY-GAADRNLQEA--QLLNGKKAQWATGLKYDANNIYLAANYGETRNATPIIN 264

QY 55 KYDGSYEAFTKKEEYKNSRFATIGTYSLNDALTFRAGLAYDKAASKTHLSASIPDPTD 114  
DB 265 GIVGAY-GAADRNLQEA--QLLNGKKAQWATGLKYDANNIYLAANYGETRNATPIIN 264





## RESULT 38

vitamin B12 receptor VC0156 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: A82358  
R:Reidberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, P.  
L.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833; PMID:10952301  
A:Accession: A82358  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-611 <HEI>  
A:Cross-references: GB:AE001405; GB:AE003852; NID:99654551; PIDN:AAF93332.1.; GSPDB:GN001  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC0156  
A:Map position: 1  
C:Superfamily: vitamin B12 receptor; tonB-dependent receptor amino-terminal homology; to

Query Match 8.9%; Score 85.5; DB 2; Length 611;  
Best Local Similarity 19.9%; Pred. No. 5.6;  
Matches 51; Conservative 23; Mismatches 97; Indels 85; Gaps 8;  
QY 1 QHNGVLGPYIGKSLTLKLPAY--WELSGFHOLDQWAIHYSKYTE-WSRFEKELRGK-- 55  
DB 189 QMNVALTESDKYGVNRPVGVNDGRHGR--SDNALLGVVHGFDESWSLFANARAYEN 246  
QY 56 -YQDGSYEAFTKKEEYKDNSRFAIGTYS----- 84  
DB 247 IYQYDINSYGRDYKEAEKDDLSFTIGTQYQSERWVSELOLTQKQSWDYTSKGYSDT 306  
QY 85 -----LNDALTLRAGLAY-----DKAASKTH----- 105  
DB 307 SDNLEQRNIQWNLVNDVNTFAGGVDWRDESYIDTADKEFORSNTAAFAVVAEWMQ 366  
QY 106 --LSASIPDTRMYY-----SIGATYKFTPNLSVDVGVFAHLRGKKKHFVETQNIKGLLL 157  
DB 367 WLEASLRFDNQEQYSGQTHNIALGQYFPEFCVKASYGSFAKPNLYQYDPSYGNV 426  
QY 158 VEADYTKATANLYGL 173  
DB 427 LQPEDADSALSFYGL 442

## RESULT 39

penicillin-binding protein 2 [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: G97053  
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cid  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: G97053  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-916 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK79218.1; PID:g15024172; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC1246

Query Match 8.9%; Score 85.5; DB 2; Length 916;  
Best Local Similarity 25.6%; Pred. No. 9.4;  
Matches 31; Conservative 18; Mismatches 33; Indels 39; Gaps 5;  
QY 28 FHQLTDQWAIHYSY-KYTEWSRFEKELR-GKYQDGSYEAFTK-----KEEYKDNS 75

DB 606 FGQVNTYTIKNSFADEGQWIAEKLOEGKYLKGSVNLTYTEDSDSKKLSTLKASIKENI 665  
QY 76 RFAL-----GTTYSLNDALTLRAGLAYDKAASKTHLSASI 110  
DB 666 RTALSTGKTDRNKTEELVTLIDTPLYKGTFSKSDVSALASGLIVYDEADEYT--QATV 723  
QY 111 P 111  
DB 724 P 724

## RESULT 40

F84148  
hypothetical protein BH3990 [Imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: F84148  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a  
A:Reference number: A83650; MUID:20512582; PMID:11058132  
A:Accession: F84148  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-221 <STO>  
A:Cross-references: GB:AP001520; GB:BA000004; NID:g10176401; PIDN:BAB07709.1; GSPDB:G  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH3990

Query Match 8.9%; Score 85; DB 2; Length 221;  
Best Local Similarity 24.1%; Pred. No. 1.7;  
Matches 34; Conservative 20; Mismatches 51; Indels 36; Gaps 6;  
QY 26 SGFHOLDQWAIHYSKYKTEWSRFEKELRGKYQ-----DSCGYEAF 65  
DB 67 SGFTMTSTSSSSSKAOKHYNKSKS-FSHAKGGYSVIDVKRSKPGFWRSLYQHYDGYGTAY 125  
QY 66 TKKEEYKDNSRFA-----IGTYSLN-----DALTLRAGLAYDKA--ASKTHLSASIP 111  
DB 126 TAKKXYKPSNSAKVATKICPNVNI SLRNSDWSLLTSRTVATDSAKYSRKYAQANFD 185  
QY 112 DTRMWSY-IGATYKFTPNLS 131  
DB 186 DLRIQWLSNLAVSEKVESKLS 206

Search completed: May 12, 2003, 09:41:39  
Job time: 20 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 09:38:34 ; Search time 13 seconds  
(without alignments)  
571.097 Million cell updates/sec

Title: US-09-995-493-52

Perfect score: 957

Sequence: 1 QHNGVLGPYICKGSLKLP.....ADYTKAVANLYGLNLNRYF 179

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	549	57.4	459	OM11_HAEIN	P43838 haemophilus
2	549	57.4	459	OM12_HAEIN	P10641 haemophilus
3	539.5	56.4	428	OM47_PASMU	P80603 pasteurella
4	332	34.7	448	FADL_ECOLI	P10384 escherichia
5	90.5	9.5	362	OMPD_SALTY	P37592 salmonella
6	90.5	9.5	875	FQCD_ECOLI	P46009 escherichia
7	88.5	9.2	522	MSAB_NEIMB	O9kin8 n peptide m
8	88	9.2	479	Y865_TREPA	O83837 treponema p
9	87.5	9.1	810	HPUB_NEIMC	P96949 neisseria m
10	86.5	9.0	522	MSAB_NEIGO	P14930 n peptide m
11	85.5	8.9	362	OMPF_ECOLI	P02931 escherichia
12	84	8.8	810	HPUB_NEIMA	O9jwa2 neisseria m
13	83.5	8.7	362	ARYH_RAT	Q02589 rattus norv
14	83	8.7	937	CS32_ECOLI	P15484 escherichia
15	82.5	8.6	522	MSAB_NEIMA	O9jwm8 n peptide m
16	81.5	8.5	362	ARYH_MOUSE	P54923 mus musculus
17	80.5	8.4	386	ALR_CLOPE	O8xm22 clostridium
18	80	8.4	265	ARC2_PHAVU	P19330 phaseolus v
19	80	8.4	365	NMPC_ECOLI	P21420 escherichia
20	80	8.4	698	TRFE_HUMAN	P02787 homo sapien
21	80	8.4	968	CTDI_HUMAN	O60716 homo sapien
22	79.5	8.3	265	ARCL_PHAVU	P19329 phaseolus v
23	79	8.3	911	CTDI_MOUSE	P30999 mus musculus
24	78.5	8.2	185	PAGC_SALTY	P23988 salmonella
25	78.5	8.2	345	YE6A_HAEIN	O86241 haemophilus
26	78.5	8.2	720	FPTA_PSEAE	P42512 pseudomonas
27	78.5	8.2	1715	NX2A_RAT	O63374 rattus norv
28	78	8.2	178	AIL_YEREN	P16454 yersinia en
29	77	8.0	252	FENR_BUCAI	P57641 buchnera ap
30	77	8.0	365	PORI_BPPA2	P07238 bacterioph
31	76.5	8.0	774	FECA_ECOLI	P13036 escherichia
32	76	7.9	289	PORI_RHOBL	P39767 rhodospseudo
33	75.5	7.9	357	ARRY_HUMAN	P54922 homo sapien

34 75.5 7.9 367 1 OMPC\_ECOLI  
35 75.5 7.9 378 1 OMPC\_SALTY  
36 75 7.8 247 1 OPAG\_NEIGO  
37 75 7.8 287 1 ISPE\_THETN  
38 75 7.8 1104 1 SYV\_YEAST  
39 75 7.8 3099 1 POLG\_PEMVM  
40 74.5 7.8 239 1 OPAA\_NEIGO  
41 74.5 7.8 378 1 PT16\_MOUSE  
42 74.5 7.8 612 1 PEX5\_YEAST  
43 74.5 7.8 660 1 ALTA\_STRPN  
44 74.5 7.8 1592 1 GTF2\_STRDO  
45 74.5 7.8 1712 1 NX2A\_HUMAN

## ALIGNMENTS

RESULT 1  
OM11\_HAEIN  
ID OM11\_HAEIN STANDARD; PRT; 459 AA.  
AC P43838;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Outer membrane protein P1 precursor (OMP P1).  
GN OMP1 OR HI0401.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Rd / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA Kerlavage A.R., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";  
RL Science 269:496-512(1995).  
RN [2]  
RP IDENTIFICATION BY MASS SPECTROMETRY.  
RX MEDLINE=20137488; PubMed=10675023;  
RA Langan H., Takacs B., Evers S., Berndt P., Lahm H.W., Wlplf B.,  
RA Gray C., Fountoulakis M.;  
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";  
RL Electrophoresis 21:411-429(2000).  
CC -|- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.  
CC -|- SIMILARITY: BELONGS TO THE OMP1/FADL FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: U32723; AAC22060.1; --  
DR TIGR: HI0401; --  
KW Outer membrane; Signal; Transmembrane; Complete proteome.  
FT SIGNAL 1 22 BY SIMILARITY.  
FT CHAIN 23 459 OUTER MEMBRANE PROTEIN P1.  
SQ SEQUENCE 459 AA; 49477 MW; 43BDC98E5A39366F CRC64;  
-----

Query Match 57.4%; Score 549; DB 1; Length 459;  
Best Local Similarity 61.6%; Pred. No. 1.le-43;







CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE002363; AAF40515.1; --  
 DR HSSP: P34149; LFVG.

DR TIGR: NMB0044; --  
 DR InterPro: IPR002579; DUF25.

DR InterPro: IPR002569; PMSR.  
 DR InterPro: IPR000063; ThioRed.

DR Pfam: PF01641; DUF25; 1.  
 DR Pfam: PF01625; PMSR; 1.

DR ProDom: PD003489; PMSR; 1.  
 DR ProDom: PD004057; DUF25; 1.

DR TIGRFAMs: TIGR00357; DUF25; 1.  
 DR TIGRFAMs: TIGR00401; msta; 1.

KW Oxidoreductase; Redox-active center; Electron transport;  
 KW Multifunctional enzyme; Complete proteome.

FT DOMAIN 38 171 THIOREDOXIN.  
 FT DOMAIN 199 354 PEPTIDE METHIONINE SULFOXIDE REDUCTASE A.

FT DOMAIN 383 506 PEPTIDE METHIONINE SULFOXIDE REDUCTASE B.  
 FT ACT\_SITE 207 207 BY SIMILARITY.

FT DISULFID 68 71 REDOX-ACTIVE (BY SIMILARITY).  
 FT DISULFID 440 495 BY SIMILARITY.

FT SEQUENCE 522 AA; 58015 MW; 535A823EDB76BFD1 CRC64;  
 SQ

Query Match 9.2%; Score 88.5; DB 1; Length 522;  
 Best Local Similarity 25.0%; Pred. No. 0.89;

Matches 42; Conservative 22; Mismatches 55; Indels 49; Gaps 8;  
 QY 11 GKQ--SLTLKLPAYMELSGP-----HQLTDQWAIHYSYKTEWSRFK----- 50

DB 370 GKGFDAATYKKPSDAELKRLTLEQYQVTONSATFATFSEYDHLKFGPIYDVVSGEPL 429

QY 51 -ELRGKYQDGSYGVEATFKKEEYKDNRSFAIGTTYSNLDAITLRLAGLAYD-----KAAS 102

DB 430 FSSADKYDSCGWPSETRPIDAK-----SVTEHDDFSYNNMRRTFVRSHAA 474

QY 103 KTHLSASIPDTR----MWYSI-CATYKFTPNLSVD-VGFAHLRGKKK 144

DB 475 DSHLGHVFPDGPDKGLRYCINGASLKFIPLEQMDAAGYALKGKVK 522

RESULT 8  
 Y865\_TREPA STANDARD; PRT; 479 AA.

ID Y865\_TREPA STANDARD; PRT; 479 AA.  
 AC O83837;

DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein TP0865 precursor.

GN TP0865.  
 OS Treponema pallidum.

OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
 OX NCBI\_TaxID=160;

RN [1]  
 RC SEQUENCE FROM N.A.

RP STRAIN=Nichols;  
 RX MEDLINE=98332770; PubMed=9655876;

RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,  
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,

RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,

RA McDonald L., Artisch P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
 RA Hatcher B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,

RA Venter J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 RT spirochete."

RL Science 281:375-388(1998).  
 CC -!- SIMILARITY: BELONGS TO THE UPF0164 FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE001256; AAC65833.1; --  
 DR TIGR: TP0865; --

DR InterPro: IPR005362; UPF0164.  
 DR Pfam: PF03687; UPF0164; 1.

KW Hypothetical protein; Signal; Complete proteome.  
 FT SIGNAL 1 49 POTENTIAL.

FT CHAIN 50 479 HYPOTHETICAL PROTEIN TP0865.  
 SQ SEQUENCE 479 AA; 52438 MW; 67592F18F1326955 CRC64;

Query Match 9.2%; Score 88; DB 1; Length 479;  
 Best Local Similarity 25.2%; Pred. No. 0.89;

Matches 40; Conservative 22; Mismatches 59; Indels 38; Gaps 8;  
 QY 29 HQLTDQWAIHYSYKTEWSRFKELRGKYQDGSYGVEATFKKEEYKDNRSFAIGTTYSL--- 85

DB 301 HATNSSFFILACAYQPIRWFLF-----GTGIEMKYNVOEFADNNRFRYGVAFLLPV 351

QY 86 -----NDALTLRAGLAYDKAASKTHLSASIPDTRMYSIGATYKFTPNLSVDV--- 134

DB 352 QYVAFSGSNVELT---GLASDIRA-----SAGV-EFKSTWVRVDLTYYESDKDEHVISCG 402

QY 135 --GFAHLRGKKKHF---VETQNIKGLLLVADYTTTKATA 168

DB 403 IAGFFN-RDRRKLEKEVYTYLRLGRLHYDAQHYEAI 440

RESULT 9  
 HPUB\_NEIMC STANDARD; PRT; 810 AA.

ID HPUB\_NEIMC STANDARD; PRT; 810 AA.  
 AC P96949;

DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hemoglobin-haptoglobin utilization protein B precursor.

GN HPUB.  
 OS Neisseria meningitidis (serogroup C).

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=135720;

RN [1]  
 RC SEQUENCE FROM N.A.

RP STRAIN=DNM2 / Serogroup C / Serotype 2a;  
 RX MEDLINE=97206152; PubMed=9157245;

RA Lewis L.A., Gray E., Wang Y.-P., Roe B.A., Dyer D.W.;  
 RT "Molecular characterization of hpuAB, the haemoglobin-haptoglobin-

RT utilization operon of Neisseria meningitidis";  
 RL Mol. Microbiol. 23:737-749(1997).

CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE  
 CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX AND IS REQUIRED FOR HEME UPTAKE.

CC -!- SUBCELLULAR LOCATION: Outer membrane (potential).  
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U73112; AAC44893.2; --  
 DR InterPro: IPR000531; TonB\_boxC.

DR Pfam: PF00593; TonB\_boxC; 1.  
 DR PROSITE: PS00430; TONB\_DEPENDENT\_REC\_1; FALSE\_NEG.

DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 KW Outer membrane; Transport; TonB box; Signal; Receptor.



OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=83090452; PubMed=6294623;  
 RA Inokuchi K., Mutoh N., Matsuyama S.-I., Mizushima S.;  
 RT "Primary structure of the ompF gene that codes for a major outer  
 membrane protein of Escherichia coli K-12";  
 RN Nucleic Acids Res. 10:6957-6968(1982).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12";  
 RN Science 277:1453-1474(1997).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 Kimura S., Kitagawa K., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 Yano M., Horiuchi T.;  
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 12.7-28.0 min region on the linkage map";  
 RN DNA Res. 3:137-155(1996).  
 [4]  
 RP SEQUENCE OF 1-37 FROM N.A.  
 RX MEDLINE=82139379; PubMed=7037455;  
 RA Mutoh N., Inokuchi K., Mizushima S.-I.;  
 RT "Amino acid sequence of the signal peptide of OmpF, a major outer  
 membrane protein of Escherichia coli";  
 RN FEBS Lett. 137:171-174(1982).  
 [5]  
 RP SEQUENCE OF 23-362.  
 RX MEDLINE=82256494; PubMed=7049161;  
 RA Chen R., Kramer C., Schmidmayr W., Chen-Schmeisser U., Henning U.;  
 RT "Primary structure of major outer-membrane protein I (ompF protein,  
 porin) of Escherichia coli B/r";  
 RN Biochem. J. 203:33-43(1982).  
 [6]  
 RP SEQUENCE OF 33-63 FROM N.A.  
 RX MEDLINE=86033642; PubMed=2997131;  
 RA Nogami T., Mizuno T., Mizushima S.;  
 RT "Construction of a series of ompF-ompC chimeric genes by in vivo  
 homologous recombination in Escherichia coli and characterization of  
 the translational products";  
 RN J. Bacteriol. 164:797-801(1985).  
 [7]  
 RP SEQUENCE OF 23-34 AND 39-47.  
 RC STRAIN=K12 / EMG2;  
 RX MEDLINE=97443975; PubMed=9298646;  
 RA Link A.J., Robison K., Church G.M.;  
 RT "Comparing the predicted and observed properties of proteins encoded  
 in the genome of Escherichia coli K-12";  
 RN Electrophoresis 18:1259-1313(1997).  
 [8]  
 RP SEQUENCE OF 23-27.  
 RC STRAIN=K12 / W3110;  
 RX MEDLINE=98291876; PubMed=9629924;  
 RA Molloy M.P., Herbert B.R., Walsh B.J., Tyler M.I., Traini M.,  
 Sanchez J.-C., Hochstrasser D.F., Williams K.L., Gooley A.A.;  
 RT "Extraction of membrane proteins by differential solubilization for  
 separation using two-dimensional gel electrophoresis";

RL Electrophoresis 19:837-844(1998).  
 [9]  
 RN X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
 RX MEDLINE=92375189; PubMed=1380671;  
 RA Cowan S.W., Schirmer T., Rummel G., Steiert M., Ghosh R.,  
 Paupit R.A., Jansonius J.N., Rosenbusch J.P.;  
 RT "Crystal structures explain functional properties of two E. coli  
 porins";  
 RN Nature 358:727-733(1992).  
 [10]  
 RN X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF MUTANT ASP-141.  
 RX MEDLINE=95024177; PubMed=7524100;  
 RA Jeanteur D., Schirmer T., Fourel D., Simonet V., Rummel G., Widmer C.,  
 Rosenbusch J.P., Pattus F., Pages J.M.;  
 RT "Structural and functional alterations of a colicin-resistant mutant  
 of OmpF porin from Escherichia coli";  
 RN Proc. Natl. Acad. Sci. U.S.A. 91:10675-10679(1994).  
 [11]  
 RN X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).  
 RX MEDLINE=99060042; PubMed=9843370;  
 RA Phale P.S., Philippsen A., Kiefhaber T., Koebnik R., Phale V.P.,  
 Schirmer T., Rosenbusch J.P.;  
 RT "Stability of trimeric OmpF porin: the contributions of the latching  
 loop L2";  
 RN Biochemistry 37:15663-15670(1998).  
 CC -1- FUNCTION: OMPF IS A PORIN THAT PASSIVE DIFFUSION PORES WHICH ALLOW  
 SMALL MOLECULAR WEIGHT HYDROPHILIC MATERIALS ACROSS THE OUTER  
 MEMBRANE. IT IS ALSO RECEPTOR FOR THE BACTERIOPHAGE T2.  
 CC -1- SUBUNIT: HOMOTRIMER.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.  
 CC -1- SIMILARITY: BELONGS TO THE OMP/PHOE FAMILY OF PORINS.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; J01655; AAA24244.1; -;  
 DR EMBL; AE000195; AAC74015.1; -;  
 DR EMBL; D90730; BAA35675.1; -;  
 DR EMBL; D90731; BAA35681.1; -;  
 DR PIR; A03431; MMECF.  
 DR PIR; A25029; A25029.  
 DR PDB; 2OMP; 07-DEC-95.  
 DR PDB; 1OPF; 07-FEB-95.  
 DR PDB; 1GFN; 07-DEC-96.  
 DR PDB; 1GFN; 07-DEC-96.  
 DR PDB; 1GFN; 07-DEC-96.  
 DR PDB; 1GFP; 07-DEC-96.  
 DR PDB; 1GFO; 07-DEC-96.  
 DR PDB; 1MPF; 07-FEB-95.  
 DR PDB; 1BT9; 13-JAN-99.  
 DR SWISS-2DPAGE; P02931; COLI.  
 DR ECO2DBASE; B036.0; 6TH EDITION.  
 DR EcoGene; EG10671; ompF.  
 DR InterPro; IPR003225; OMP\_2.  
 DR InterPro; IPR001702; Porin\_gram-ve.  
 DR Pfam; PF00267; Gram-ve\_porins; 1.  
 DR PRINTS; PR00182; ECOLNEIPORIN.  
 DR PROSITE; PS00576; GRAM\_NEG\_PORIN; 1.  
 KW Outer membrane; Transmembrane; Porin; Phage recognition; Signal;  
 3D-structure; Complete proteome.  
 FT SIGNAL 1 22  
 FT CHAIN 23 362 OUTER MEMBRANE PROTEIN F.  
 FT CONFLICT 88 Q -> E (IN REF. 5).  
 FT CONFLICT 139 139 E -> G (IN REF. 5).  
 FT CONFLICT 284 284 Q -> L (IN REF. 5).  
 FT STRAND 24 28  
 FT TURN 29 30  
 FT STRAND 31 45

```

FT TURN 50 51
FT STRAND 53 53
FT STRAND 58 59
FT STRAND 62 72
FT STRAND 77 88
FT TURN 95 100
FT STRAND 102 112
FT TURN 113 115
FT STRAND 116 124
FT TURN 126 127
FT HELIX 128 131
FT TURN 132 132
FT TURN 143 144
FT TURN 147 148
FT TURN 150 151
FT STRAND 154 163
FT HELIX 165 169
FT TURN 171 172
FT STRAND 173 180
FT STRAND 183 183
FT TURN 188 190
FT STRAND 192 192
FT STRAND 195 204
FT TURN 205 206
FT STRAND 207 217
FT HELIX 220 223
FT TURN 224 224
FT STRAND 227 227
FT STRAND 232 244
FT TURN 245 246
FT STRAND 247 257
FT STRAND 261 264
FT TURN 265 268
FT STRAND 269 272
FT STRAND 275 285
FT TURN 288 289
FT STRAND 291 305
FT TURN 306 308
FT STRAND 309 324
FT STRAND 329 338
FT STRAND 351 351
FT STRAND 353 361
SQ SEQUENCE 362 AA; 39333 MW; 3F0974D96DB65464 CRC64;

Query Match 8.9%; Score 85.5; DB 1; Length 362;
Best Local Similarity 25.2%; Pred. No. 1.1;
Matches 36; Conservative 22; Mismatches 60; Indels 25; Gaps 6;

QY 4 CVLGPYIGKSLTLKLPAYWELSGFHQLTDOWA-----IHSYKYTEWSRFKELRG 54
DQ 208 GIVGAY-GAADRITNLQEA--QPLNGKKAQWATGLKYDANNIYLAANYGETRNATPITN 264
QY 55 KYQDGSVEAFKTEKYKDNRSFAIGTYSLNDALTLRAGLAYDKAASKTHLSASIPDPTD 114
DQ 265 KFTWTSFANKTQVLLVAQYQDFEG-----LRPSIATYKSKAKD--VEGIGDVD 312
QY 115 RM-WYSIGATYKFTPNLSVDVGF 136
DQ 313 LVNYFEVGATYFKNKMSYVDY 335

RESULT 12
HPUB_NEIMA
ID HPUB_NEIMA STANDARD; PRT; 810 AA.
AC Q9JWA2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemoglobin-haptoglobin utilization protein B precursor.
GN HPUB OR NMA0474.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OC NCBI_TaxID=65699;

RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=22491 / Serogroup A / Setotype 4A;
RA MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Davis K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RA "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
CC -1- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX AND IS REQUIRED FOR HEME UPTAKE (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL162753; CAB83769.1; -
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC; 1.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Transport; TonB box; Signal; Receptor;
KW Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 810 HEMOGLOBIN-HAPTOGLOBIN UTILIZATION
FT PROTEIN B.
FT SITE 793 810 TONB C-TERMINAL BOX.
SQ SEQUENCE 810 AA; 90570 MW; D38DE1DCA3CA5A6E CRC64;

Query Match 8.8%; Score 84; DB 1; Length 810;
Best Local Similarity 24.2%; Pred. No. 4;
Matches 47; Conservative 17; Mismatches 74; Indels 56; Gaps 10;

QY 22 YWELSGFHQLTDOWAIHY-----SYKYTEWSRFKELRGKYQD--GSGYEAFTK----- 67
DQ 378 YHSFRHRIQNTAQTADFQKOLFQKAVAAQYGLGGKGDNANSYFAKLYDPKILA 437
QY 68 -----KEEYKD---NSRFAIGTYSLNDALTLRAGLAYDKAASK-----T 104
DQ 438 SNQAKITMLIENRSKYRFAYWNAFHILG-----NDRFLNAGIRYDKNSSAKDPPKVT 493
QY 105 HLSASIP-DTDR---MWYSIGATYKFTPNLSVDVGFAGHLRGKK--HFVTONIKGLL 156
DQ 494 AIRQIIPHLSGERAHAGFSYGTGDFWRFTKHL-----HLAKYSTGFRAPTSDETWLL 546
QY 157 LVEADYTTKATANL 170
DQ 547 FPHDPFVLKANPNL 560

RESULT 13
ARHY_RAT
ID ARHY_RAT STANDARD; PRT; 362 AA.
AC Q02589;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADP-ribosylarginine hydrolase (EC 3.2.2.19) (ADP-ribose-L-arginine
DE cleaving enzyme).
GN ADPRH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```





RA Whitehead S., Spratt B.G., Barrell B.G.;  
 RT "Complete DNA sequence of a serogroup A strain of *Neisseria*  
 RL meningitidis 22491";  
 RN Nature 404:502-506(2000).  
 RP [2]  
 RP IDENTIFICATION OF THE METHIONINE SULFOXIDE REDUCTASE ACTIVITIES (MSRA  
 AND MSRB).  
 RC STRAIN=22491 / Serogroup A / Serotype 4A;  
 RA Orly A., Boschi-Muller S., Marraud M., Sanglier-Cianferani S.,  
 van Dorsselaar A., Brabant G.;  
 RA "Characterization of the methionine sulfoxide reductase activities  
 RT of PflB, a probable virulence factor from *Neisseria meningitidis*.";  
 RL J. Biol. Chem. 277:10000-10006(2002).  
 CC -1- FUNCTION: Has an important function as a repair enzyme for  
 CC proteins that have been inactivated by oxidation (By similarity).  
 CC Catalyzes the reversible oxidation-reduction of methionine  
 CC sulfoxide in proteins to methionine.  
 CC -1- CATALYTIC ACTIVITY: Protein L-methionine + oxidized thioredoxin =  
 CC protein L-methionine S-oxide + reduced thioredoxin.  
 CC -1- DOMAIN: Possesses 2 methionine sulfoxide reductase domains (A/MSrA  
 CC and B/MSrB) and 1 N-terminal thioredoxin domain. The domain B  
 CC exhibits a thioredoxin dependent methionine sulfoxide reductase  
 CC activity; the Cys-495 is probably involved in the reduction of  
 CC MetSO and in formation of the sulfenic acid derivative. The  
 CC regeneration of Cys-495 is probably done via formation of a  
 CC disulfide bond with Cys-440 followed by its reduction by  
 CC thioredoxin.  
 CC -1- MISCELLANEOUS: The domain msrB is stereospecific for the R isomer  
 CC of the sulfoxide of MetSO whereas the domain msrA is  
 CC stereospecific for the S isomer.  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE THIOREDOXIN  
 CC FAMILY.  
 CC -1- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE MSRA MET  
 CC SULFOXIDE REDUCTASE FAMILY.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE MSRB MET  
 CC SULFOXIDE REDUCTASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AL162752; CAB83597.1;  
 CC HSSP; P54149; IPVG.  
 CC InterPro; IPR002579; DUF25.  
 CC InterPro; IPR002569; PMSR.  
 CC InterPro; IPR000063; ThioRed.  
 CC Pfam; PF01641; DUF25; 1.  
 CC Pfam; PF01625; PMSR; 1.  
 CC ProDom; PD003489; PMSR; 1.  
 CC ProDom; PD004057; DUF25; 1.  
 CC TIGRFAMs; TIGR00357; DUF25; 1.  
 CC TIGRFAMs; TIGR00401; msrA; 1.  
 CC Oxidoreductase; Redox-active center; Electron transport;  
 CC Multifunctional enzyme; Complete proteome.  
 CC FT DOMAIN 38 171 THIOREDOXIN.  
 CC FT DOMAIN 199 354 PEPTIDE METHIONINE SULFOXIDE REDUCTASE A.  
 CC FT DOMAIN 383 506 PEPTIDE METHIONINE SULFOXIDE REDUCTASE B.  
 CC FT ACT\_SITE 207 207 BY SIMILARITY.  
 CC FT DISULFID 68 71 REDOX-ACTIVE (BY SIMILARITY).  
 CC FT DISULFID 440 495 PROBABLE.  
 CC SEQUENCE 522 AA; 58015 MW; F61E8EA7189F0667 CRC64;

Query Match 8.6%; Score 82.5; DB 1; Length 522;  
 Best Local Similarity 24.4%; Pred. No. 3.2;  
 Matches 41; Conservative 22; Mismatches 56; Indels 49; Gaps 8;  
 QY 11 GKGLMEY-LPEAKYITQSGY--HOLDTQWAIHYKYKTEWSRFFK-----50  
 DB 370 GKGFDAATYKPSDAELKRLTTEQYQVTONSATEYAFSHEYDLFRPGIYVDVWSGEPL 429

QY 51 -ELRGKYDGSYGAEFTKKEEYKONSRAFTGTTYSLNDALTLRAGLAYD-----KAAS 102  
 DB 430 FSSADKYDSCGWPSTFRPIDAK-----SVTEHDDPSYNNRRTEVRSAA 474  
 QY 103 KTHLSASIPDTR----MWYSI-GATYKFTPNLSVD-VGFAHLRGKKK 144  
 DB 475 DSHLGHVDPDPRDKGLRYCINGASLKFTPLEQMDAAGYALKSKVK 522  
 RESULT 16  
 ARMY\_MOUSE  
 ID ARMY\_MOUSE STANDARD; PRT; 362 AA.  
 AC P54923;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE ADP-ribosylarginine hydrolase (EC 3.2.2.19) (ADP-ribose-L-arginine  
 DE cleaving enzyme).  
 GN ADPRH.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93352593; PubMed=8349667;  
 RA Takada T., Iida K., Moss J.;  
 RT "Cloning and site-directed mutagenesis of human ADP-ribosylarginine  
 RT hydrolase.";  
 RL J. Biol. Chem. 268:17837-17843(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Aoki K., Shoemaker M., Moss J.;  
 RT "Genomic organization and promoter analysis of the mouse  
 RT ADP-ribosylarginine hydrolase gene.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CATALYZES REVERSE REACTION OF MONO-ADP-RIBOSYLATION.  
 CC -1- CATALYTIC ACTIVITY: N(2)-(ADP-D-ribosyl)-L-arginine + H(2)O = L-  
 CC arginine + ADP-ribose.  
 CC -1- ENZYME REGULATION: SYNERGISTICALLY STIMULATED BY MAGNESIUM AND  
 CC DITHIOERITOL (DTT) IN VITRO.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; LI3290; AAA37259.1;  
 CC DR EMBL; AF244347; AAF86223.1;  
 CC DR EMBL; BC003437; AAH03437.1;  
 CC DR MGD; MGI:1098234; Adprh.  
 CC KW Hydrolase; Magnesium  
 CC SQ SEQUENCE 362 AA; 40068 MW; E06BCAEF69BA7C4E CRC64;  
 Query Match 8.5%; Score 81.5; DB 1; Length 362;  
 Best Local Similarity 27.2%; Pred. No. 2.6;  
 Matches 37; Conservative 18; Mismatches 54; Indels 27; Gaps 8;  
 QY 11 GKSLTLKLP---AYWELSGFHQLTDQWAIHYKYKTEWSRFRKELRGKYDGSYGAEFTK 67  
 DB 197 GKGLMEY-LPEAKYITQSGY--FVKENLQHWSEYFEWEKYLELRG-ILDGNSAPVFPQ 252  
 QY 68 KEEYKONSRAFTGTTYS-----LNDALTLRAGLAYDKAASKTHLSASIPD 114  
 DB 253 PFGVKERDQYIDVSYSCWCGSSGHDPAMAYDAL-LAAGDSWKELAIRAFFIGGSDST 311





RT "A cloned gene for human transferrin.";  
 RL Ann. N.Y. Acad. Sci. 646:140-154(1991).  
 RN [4]  
 RP SEQUENCE FROM N.A., AND VARIANT ATRANSFERRINEMIA PRO-477.  
 RX MEDLINE-20563920; PubMed-11110675;  
 RA Beutler E., Gelbart T., Lee P.L., Trevino R., Fernandez M.A.,  
 RA Fairbanks V.F.;  
 RT "Molecular characterization of a case of atransferrinemia.";  
 RL Blood 96:4071-4074(2000).  
 RN [5]  
 RP SEQUENCE OF 99-698 FROM N.A.  
 RC TISSUE=Fetal liver;  
 RA Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Bi J., Zhang Y., Liu M.,  
 RA He F.;  
 RT "Functional prediction of the coding sequences of 33 new genes deduced  
 by analysis of cDNA clones from human fetal liver.";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE OF 422-698 FROM N.A.  
 RX MEDLINE-84153910; PubMed-6322780;  
 RA Uzan G., Frain M., Park I., Besmond C., Maessen G., Trepap J.S.,  
 RA Zakim M.M., Kahn A.;  
 RT "Molecular cloning and sequence analysis of cDNA for human  
 transferrin.";  
 RL Biochem. Biophys. Res. Commun. 119:273-281(1984).  
 RN [7]  
 RP SEQUENCE OF 20-698.  
 RX MEDLINE-83160878; PubMed-6833213;  
 RA McGillivray R.T.A., Mendez E., Shewale J.G., Sinha S.K.,  
 RA Lineback-Zins J., Brew K.;  
 RT "The primary structure of human serum transferrin. The structures of  
 seven cyanogen bromide fragments and the assembly of the complete  
 structure.";  
 RL J. Biol. Chem. 258:3543-3553(1983).  
 RN [8]  
 RP SEQUENCE OF 73-698 FROM N.A.  
 RX MEDLINE-85216459; PubMed-3858812;  
 RA Park I., Schaeffer E., Sidoli A., Baralle F.E., Cohen G.N.,  
 RA Zakim M.M.;  
 RT "Organization of the human transferrin gene: direct evidence that it  
 originated by gene duplication.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:3149-3153(1985).  
 RN [9]  
 RP SEQUENCE OF 1-14 FROM N.A.  
 RX MEDLINE-87066744; PubMed-3786138;  
 RA Lucero M.A., Schaeffer E., Cohen G.N., Zakim M.M.;  
 RT "The 5' region of the human transferrin gene: structure and potential  
 regulatory sites.";  
 RL Nucleic Acids Res. 14:8692-8692(1986).  
 RN [10]  
 RP SEQUENCE OF 1-72 AND 291-300 FROM N.A.  
 RX MEDLINE-87192006; PubMed-3106157;  
 RA Adrian G.S., Korinek B.W., Bowman B.H., Yang F.;  
 RT "The human transferrin gene: 5' region contains conserved sequences  
 which match the control elements regulated by heavy metals,  
 glucocorticoids and acute phase reaction.";  
 RL Gene 49:167-175(1986).  
 RN [11]  
 RP SEQUENCE OF 45-72 FROM N.A.  
 RX PubMed-10931525;  
 RA de Arriba Zerpa G.A., Saleh M.C., Fernandez P.M., Guillou F.,  
 RA Espinosa de los Monteros A., de Vellis J., Zakim M.M., Baron B.;  
 RT "Alternative splicing prevents transferrin secretion during  
 differentiation of a human oligodendrocyte cell line.";  
 RL J. Neurosci. Res. 61:388-395(2000).  
 RN [12]  
 RP SEQUENCE OF 564-624 FROM N.A., AND VARIANT TP\*C2.  
 RC TISSUE=Brain;  
 RX MEDLINE-97418135; PubMed-9272172;  
 RA Namekata K., Oyama F., Imagawa M., Ihara Y.;  
 RT "Human transferrin (Tf): a single mutation at codon 570 determines Tf  
 C1 or Tf C2 variant.";  
 RL Hum. Genet. 100:457-458(1997).

RN [13]  
 RP SEQUENCE OF 564-624 FROM N.A.  
 RA Tsuchida S., Ikemoto S., Kajii E.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [14]  
 RP SEQUENCE OF 636-696 FROM N.A.  
 RX MEDLINE-89386721; PubMed-2780570;  
 RA Duguid J.R., Bohmont C.W., Liu N.G., Tourtellotte W.W.;  
 RT "Changes in brain gene expression shared by scrapie and Alzheimer  
 disease.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:7260-7264(1989).  
 RN [15]  
 RP SEQUENCE OF 263-266; 454-458; 531-538 AND 589-595.  
 RC TISSUE=Heart;  
 RX MEDLINE-96007936; PubMed-7498159;  
 RA Kovalyov L.I., Shishkin S.S., Efimochkin A.S., Kovalyova M.A.,  
 RA Brshova E.S., Egorov T.A., Musalyanov A.K.;  
 RT "The major protein expression profile and two-dimensional protein  
 database of human heart.";  
 RL Electrophoresis 16:1160-1169(1995).  
 RN [16]  
 RP DISULFIDE BONDS  
 RX MEDLINE-82222166; PubMed-6953407;  
 RA McGillivray R.T.A., Mendez E., Sinha S.K., Sutton M.R.,  
 RA Lineback-Zins J., Brew K.;  
 RT "The complete amino acid sequence of human serum transferrin.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 79:2504-2508(1982).  
 RN [17]  
 RP MUTAGENESIS.  
 RX MEDLINE-92031336; PubMed-1932003;  
 RA Woodworth R.C., Mason A.B., Funk W.D., McGillivray R.T.A.;  
 RT "Expression and initial characterization of five site-directed  
 mutants of the N-terminal half-molecule of human transferrin.";  
 RL Biochemistry 30:10824-10829(1991).  
 RN [18]  
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 22-350.  
 RX MEDLINE-98272665; PubMed-9609685;  
 RA Macgillivray R.T.A., Moore S.A., Chen J., Anderson B.F., Baker H.,  
 RA Luo Y., Bewley M.C., Smith C.A., Murphy M.E.P., Wang Y., Mason A.B.,  
 RA Woodworth R.C., Brayer G.D., Baker E.N.;  
 RT "Two high-resolution crystal structures of the recombinant N-lobe of  
 human transferrin reveal a structural change implicated in iron  
 release.";  
 RL Biochemistry 37:7919-7928(1998).  
 RN [19]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 22-350.  
 RX MEDLINE-98434369; PubMed-9760232;  
 RA Jeffrey P.D., Bewley M.C., Macgillivray R.T.A., Mason A.B.,  
 RA Woodworth R.C., Baker E.N.;  
 RT "Ligand-induced conformational change in transferrins: crystal  
 structure of the open form of the N-terminal half-molecule of human  
 transferrin.";  
 RL Biochemistry 37:13978-13986(1998).  
 RN [20]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 23-352.  
 RX MEDLINE-99155227; PubMed-10029548;  
 RA Bewley M.C., Tam B.M., Grewal J., He S., Shewry S., Murphy M.E.P.,  
 RA Mason A.B., Woodworth R.C., Baker E.N., Macgillivray R.T.A.;  
 RT "X-ray crystallography and mass spectroscopy reveal that the N-lobe  
 of human transferrin expressed in *Pichia pastoris* is folded correctly  
 but is glycosylated on serine-32.";  
 RL Biochemistry 38:2535-2541(1999).  
 RN [21]  
 RP VARIANT SER-142.  
 RX MEDLINE-98019079; PubMed-9358047;  
 RA Evans P., Kemp J.;  
 RT "Exon/intron structure of the human transferrin receptor gene.";  
 RL Gene 199:123-131(1997).  
 RN [22]  
 RP VARIANT GLU-646.  
 RX PubMed-9803271;  
 RA Pang H., Koda Y., Soejima M., Kimura H.;  
 RT "Identification of a mutation (A1879G) of transferrin from cDNA

prepared from peripheral blood cells.":  
 [23]  
 RN Ann. Hum. Genet. 62:271-274(1998).  
 RP VARIANTS SER-277; SER-589 AND GLU-671, AND CHARACTERIZATION OF  
 RP VARIANTS SER-277.  
 RX MEDLINE=21560268; PubMed=11703331;  
 RA Lee P.L., Halloran C., Trevino R., Fellitti V., Beutler E.;  
 RT "Human transferrin G277S mutation: a risk factor for iron deficiency  
 anemia.";  
 RL Br. J. Haematol. 115:329-333(2001).  
 RN [24]  
 RP VARIANTS SER-277 AND SER-589.  
 RX MEDLINE=21558434; PubMed=11702220;  
 RA Douabin-Gicquel V., Soriano N., Ferran N., Wojcik F., Palierne E.,  
 RA Tanim S., Jovelin T., McKie A.T., Le Gall J.-Y., David V., Mosser J.;  
 RT "Identification of 96 single nucleotide polymorphisms in eight genes  
 involved in iron metabolism: efficiency of bioinformatic extraction  
 compared with a systematic sequencing approach.";  
 RL Hum. Genet. 109:393-401(2001).  
 Query Match 8.4%; Score 80; DB 1; Length 698;  
 Best Local Similarity 27.9%; Pred. No. 7.9;  
 Matches 34; Conservative 15; Mismatches 53; Indels 20; Gaps 6;  
 Qy 32 TDQNAIHVSYYKTEWSRKE-LRSGYQDSGYEAFPTKKEEYKDSRFAIGTYSUNDAIT 90  
 Db 24 TVRCAVSEHEATKQSFDRHMKSVIPSDGPSVACVKASYLDCTIR-AIAANEA--DAVT 80  
 Qy 91 LRAGLAYDKAASKTHLS-----ASIPDTRMYSIGATYKFTPNLSVDVGEA--HLRG 141  
 Db 81 LDAGLVYDAYLAPNNKLPVAAEFYKSGREDPTFYAVAVKK-----DSGFQMNQLRG 133  
 Qy 142 KK 143  
 Db 134 KK 135  
 RESULT 21  
 CTDL\_HUMAN STANDARD; PRT; 968 AA.  
 AC Q60716; Q60715; Q60713; Q60935; Q60714; Q15088; Q9UP72; Q9UP71;  
 AC Q9UP73;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Catenin delta-1 (p120 catenin) (p120(ctn)) (Cadherin-associated Src  
 DE substrate) (CAS) (p120(cas)).  
 GN CTNND1 OR KIAA0384  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 ON NCBI\_TaxID=9606;  
 RX [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC TISSUE=Fetal kidney;  
 RX MEDLINE=98317528; PubMed=9653641;  
 RA Keirsgiblick A., Bonne S., Staes K., van Hengel J., Nolllet F.,  
 RA Reynolds A., van Roy F.;  
 RT "Molecular cloning of the human p120ctn catenin gene (CTNND1):  
 RT expression of multiple alternatively spliced isoforms.";  
 RL Genomics 50:129-146(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1AC).  
 RC TISSUE=Brain;  
 RX MEDLINE=97349984; PubMed=9205841;  
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes..VII.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 4:141-150(1997).  
 CC -1- FUNCTION: EFFICIENT TYROSINE KINASE SUBSTRATE IMPLICATED BOTH IN  
 CC CELL TRANSFORMATION BY SRC AND IN LIGAND-INDUCED RECEPTOR

SIGNALING THROUGH THE EGF, PDGF, CSF-1 AND ERBB2 RECEPTORS. THE  
 ASSOCIATION OF CATENINS TO CADHERINS PRODUCES A COMPLEX WHICH IS  
 LINKED TO THE ACTIN FILAMENT NETWORK, AND WHICH SEEMS TO BE OF  
 PRIMARY IMPORTANCE FOR CADHERINS CELL-ADHESION PROPERTIES.  
 -1- SUBUNIT: BELONGS TO A MULTIPROTEIN CELL-CELL ADHESION COMPLEX THAT  
 ALSO CONTAINS E-CADHERIN, ALPHA-CATENIN, BETA-CATENIN, AND GAMMA-  
 CATENIN. BINDS TO PRESENILIN 1 C-TERMINAL FRAGMENT AND MUTUALLY  
 COMPETES FOR E-CADHERIN.  
 -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR IN PARTICULAR CELLS.  
 -1- ALTERNATIVE PRODUCTS: 32 ISOFORMS; 1ABC (SHOWN HERE), 1AB, 1AC, 1AB,  
 1BC, 1A, 1B, 1C, 1, 2ABC, 2AB, 2AC, 2BC, 2A, 2B, 2C, 2, 3ABC, 3AB,  
 3AC, 3BC, 3C, 3B, 3C, 3, 4ABC, 4AB, 4AC, 4BC, 4A, 4B, 4C AND 4;  
 MAY BE PRODUCED BY ALTERNATIVE SPLICING. IT IS UNCERTAIN IF ALL  
 THESE ISOFORMS EXIST IN VIVO.  
 -1- PTM: PHOSPHORYLATION (BY SIMILARITY).  
 -1- DISEASE: MAY CONTRIBUTE TO CELL MALIGNANCY. COMPLETE LOSS OF  
 EXPRESSION WAS OBSERVED IN APPROXIMATELY 10% OF INVASIVE DUCTAL  
 BREAST CARCINOMAS INVESTIGATED.  
 -1- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.  
 -1- SIMILARITY: CONTAINS 10 ARM REPEATS.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; AF062319; AAC39804.1;  
 EMBL; AF062323; AAC39808.1;  
 EMBL; AF062341; AAC39826.1;  
 EMBL; AF062342; AAC39827.1;  
 EMBL; AF062322; AAC39807.1;  
 EMBL; AF062326; AAC39811.1;  
 EMBL; AF062328; AAC39813.1;  
 EMBL; AF062338; AAC39823.1;  
 EMBL; AF062324; AAC39809.1;  
 EMBL; AF062327; AAC39812.1;  
 EMBL; AF062330; AAC39815.1;  
 EMBL; AF062331; AAC39816.1;  
 EMBL; AF062333; AAC39818.1;  
 EMBL; AF062334; AAC39819.1;  
 EMBL; AF062335; AAC39820.1;  
 EMBL; AF062336; AAC39821.1;  
 EMBL; AF062339; AAC39824.1;  
 EMBL; AF062340; AAC39825.1;  
 EMBL; AF062343; AAC39828.1;  
 EMBL; AF062317; AAC39802.1;  
 EMBL; AF062325; AAC39810.1;  
 EMBL; AF062332; AAC39817.1;  
 EMBL; AF062344; AAC39829.1;  
 EMBL; AF062321; AAC39806.1;  
 EMBL; AF062320; AAC39805.1;  
 EMBL; AF062337; AAC39822.1;  
 EMBL; AF062318; AAC39803.1;  
 EMBL; AB002382; BAA20838.1;  
 Genew; HGNC:2515; CTNND1.  
 MIM; 601045; -;  
 DR InterPro; IPR000225; Armadillo.  
 DR Pfam; PF00514; Armadillo\_seg; 5.  
 DR SMART; SM00185; ARM; 4.  
 DR PROSITE; PS00176; ARM\_REPEAT; 3.  
 Cytoskeleton; Structural protein; Phosphorylation; Repeat;  
 Cell adhesion; Coiled coil; Nuclear protein; Alternative splicing.  
 FT DOMAIN 10 46  
 FT REPEAT 358 395 ARM 1.  
 FT REPEAT 398 437 ARM 2.  
 FT REPEAT 441 475 ARM 3.  
 FT REPEAT 476 516 ARM 4.  
 FT REPEAT 534 573 ARM 5.  
 FT REPEAT 583 624 ARM 6.



CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX [1]  
 RN NCBI\_TaxID=10090;  
 PC SEQUENCE FROM N.A.  
 RC STRAIN=Swiss;  
 RX MEDLINE=93096477; PubMed=1334250;  
 RA Reynolds A.B., Herbert L., Cleveland J.L., Berg S.T., Gaut J.R.;  
 RT "p120, a novel substrate of protein tyrosine kinase receptors and of  
 RT p60v-src, is related to cadherin-binding factors beta-catenin,  
 RT plakoglobin and armadillo";  
 RL Oncogene 7:2439-2445(1992).  
 RN [2]  
 RP REVISIONS.  
 RA Reynolds A.B.;  
 RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: EFFICIENT TYROSINE KINASE SUBSTRATE IMPLICATED BOTH IN  
 CC CELL TRANSFORMATION BY SRC AND IN LIGAND-INDUCED RECEPTOR  
 CC SIGNALING THROUGH THE EGF, PDGF, CSF-1 AND EBB2 RECEPTORS (BY  
 CC SIMILARITY).  
 CC -1- SUBUNIT: BINDS TO THE CYTOPLASMIC DOMAIN OF THE CELL-CELL ADHESION  
 CC MOLECULE E-CADHERIN. THE ASSOCIATION OF CATELINS TO CADHERINS  
 CC PRODUCES A COMPLEX WHICH IS LINKED TO THE ACTIN FILAMENT NETWORK,  
 CC AND WHICH SEEMS TO BE OF PRIMARY IMPORTANCE FOR CADHERINS CELL-  
 CC ADHESION PROPERTIES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR IN PARTICULAR CELLS  
 CC (BY SIMILARITY).  
 CC -1- PTM: PHOSPHORYLATED BY PROTEIN-TYROSINE KINASES.  
 CC -1- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 10 ARM REPEATS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: 217804; CAA79078.1; --  
 DR PIR: S28498; S28498.  
 DR MGD: MGI:105100; Catns.  
 DR InterPro: IPR000225; Armadillo.  
 DR Pfam: PF00514; Armadillo\_seg; 5.  
 DR SMART: SM00185; ARM; 4.  
 DR Cyskoskeleton: PS01076; ARM\_REPEAT; 3.  
 KW Cytoskeleton; Structural protein; Phosphorylation; Repeat;  
 KW Cell adhesion; Coiled coil; Nuclear protein.  
 FT DOMAIN 10 46  
 FT REPEAT 358 395  
 FT REPEAT 398 437  
 FT REPEAT 441 475  
 FT REPEAT 476 516  
 FT REPEAT 534 573  
 FT REPEAT 583 624  
 FT REPEAT 647 687  
 FT REPEAT 694 733  
 FT REPEAT 734 774  
 FT REPEAT 775 820  
 FT DOMAIN 622 628  
 FT SEQUENCE 911 AA; 101731 MW; EE18C6223948DD1 CRC64;  
 Query Match 8.3%; Score 79; DB 1; Length 911;  
 Best Local Similarity 25.2%; Pred. No. 13;  
 Matches 29; Conservative 15; Mismatches 43; Indels 28; Gaps 6;  
 QY 3 NGVLGPIYKGSLSLKLPAWELSGFHQLTDQWATHV-----SYKYTENSREKELRGK 55  
 DB 187 NGGPPYVQAG-TATL-----RNHFPPDGYGRHYEDGYPGSDNPGSVSRVTRIER 240  
 QY 56 YQDG-SGYEATTKKEEKDNRFAIGTTF-----YSINDALTLRAGLAD 98  
 DB 241 YRPSMEGYRAPSRQDVYGPQVRVGGSSVDLHRFHPPEYGLEDD---QRSMGYD 292

RESULT 24  
 PAGC\_SALTY  
 ID PAGC\_SALTY STANDARD; PRT; 185 AA.  
 AC P23988;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Virulence membrane protein pagC precursor.  
 GN PAGC OR STM1246.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 14028;  
 RX MEDLINE=91100323; PubMed=1846140;  
 RA Puikinen W.S., Miller S.I.;  
 RT "A Salmonella typhimurium virulence protein is similar to a Yersinia  
 RT enterocolitica invasion protein and a bacteriophage lambda outer  
 RT membrane protein";  
 RL J. Bacteriol. 173:86-93(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McLelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 RT LT2";  
 RL Nature 413:852-856(2001)  
 CC -1- FUNCTION: ESSENTIAL FOR FULL VIRULENCE AND SURVIVAL WITHIN  
 CC MACROPHAGES.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.  
 CC -1- SIMILARITY: BELONGS TO THE AIL/OMPX/PAGC/LOM FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M55546; AAA27179.1; --  
 DR EMBL: AE008754; AAL20174.1; --  
 DR PIR: A39185; A39185.  
 DR HSP; P36546; IQJ8.  
 DR StyGene: SG10676; pagC.  
 DR InterPro: IPR000758; Enterovir\_OMP.  
 DR PROSITE: PS00694; ENT\_VIR\_OMP\_1; 1.  
 DR PROSITE: PS00695; ENT\_VIR\_OMP\_2; 1.  
 KW Outer membrane; Transmembrane; Signal; Virulence; Complete proteome.  
 FT SIGNAL 1 23 POTENTIAL.  
 FT CHAIN 24 185 VIRULENCE MEMBRANE PROTEIN PAGC.  
 FT CONFLICT 33 33 A->ARYA (IN REF. 1).  
 FT SEQUENCE 185 AA; 20183 MW; 10B4E785032ECB9E CRC64;  
 Query Match 8.2%; Score 78.5; DB 1; Length 185;  
 Best Local Similarity 23.9%; Pred. No. 2.1;  
 Matches 45; Conservative 20; Mismatches 56; Indels 67; Gaps 10;  
 QY 33 DQWAIHSYKYTEWSRKEKELRG-----KYQDGSYEATTKKEEKDNRFA--E----- 78  
 DB 24 DTNAPSVGYAQSKYQDFKIRGVNWKYREDSPV-SFISLSLYLGDRQASVGEPI 82  
 QY 79 -----IGTYSINDALTL--RAGLAYDAASKTHLSASIPDPTD----- 114  
 DB 83 HYHDKFEVKYGLVGPAYRLSDNFSLYALAGVGTVATFKEH---STQDGDSPSNKISS 139



Db	91	AQXDGKLLYGLGVYKFTPNIA--TFFNHAE	SPQNNRTLIINGELPABQGSFETGLK	148
Qy	164	-----TKATANLYGLN	174	
Db	149	YENAYLNATVALFNIN	164	

```

RESULT_26
FPTA_PSEAE
ID   FPTA_PSEAE          STANDARD;             PRT:    720 AA.
AC   P42512;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Fe(III)-pyochelin receptor precursor.
GN   FPTA OR PA4221.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
NCBI_TaxID=287;
[1]
SEQUENCE FROM N.A.O., AND SEQUENCE OF 39-51.
STRAIN=PAO / IA602;
MEDLINE=94117363; PubMed=8288523;
Ankenbauer R.G.; Quan H.N.;
"rPta, the Fe(III)-pyochelin receptor of Pseudomonas aeruginosa
phenolate siderophore receptor homologous to hydroxamate sidero-
ceptors." J. Bacteriol. 176:307-319(1994).
[2]

```

Query Match 8.2%; Score 78.5; DB 1; Length 720;  
Best Local Similarity 24.0%; Pred. No. 11;  
Matches 35; Conservative 17; Mismatches 47; Indels 47; Gaps 8;  
QY 23 WEJSGFHOLDWAIHYSYKYTEWSREKRGKYODSGYEAFTKKBEYKDNRSFAICTT 82

```

Db 577 FELEGTGLTPYSLWAGYTYTSTLEYLKD--SQNDSCGTRYSTFTPRH----- 621
Qy 83 YSLNDALTLRAGLAYDKAASKTHLSASIPDQRMWYSIGATYKTPNLSVDV----- 134
Db 622 -----LRLWSNVD-----LPWQDRRW-SVGGGLQAQSDYSVDYRGVSMRQ 662
Qy 135 GFA--HLR-GKK--KHFEVETQIKGL 155
Db 663 GYALVNMELGYKIDHWTAAVNVNVL 688

RESULT 27
NX2A_RAT
ID NX2A_RAT STANDARD; PRT; 1715 AA.
AC Q63374; Q63375;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurexin 2-alpha precursor (Neurexin II-alpha).
OS NRXN2.
GN Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A., VARIANT LEU-434, AND ALTERNATIVE SPLICING.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=92320296; PubMed=1621094;
RA Ushkaryov Y.A., Petrenko A.G., Geppert M., Suedhof T.C.;
RT "Neurexins: Synaptic cell surface proteins related to the
RL alpha-latrotoxin receptor and laminin.";
RN Science 257:50-56(1992).
[2]
SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Brain;
RX MEDLINE=95209856; PubMed=7695896;
RA Ullrich B., Ushkaryov Y.A., Suedhof T.C.;
RT "Cartography of neurexins: more than 1000 isoforms generated by
RL alternative splicing and expressed in distinct subsets of neurons.";
RN Neuron 14:497-507(1995).
[3]
INTERACTION WITH NEUREXOPHILIN 1.
RX MEDLINE=99074239; PubMed=9856994;
RA Missler M., Hammer R.E., Suedhof T.C.;
RT "Neurexophilin binding to alpha-neurexins. A single LNS domain
RL functions as an independently folding ligand-binding unit.";
RN J. Biol. Chem. 273:34716-34723(1998).
[4]
INTERACTION WITH ALPHA-DYSTROGLYCAN.
RX MEDLINE=21363578; PubMed=11470830;
RA Sugita S., Saito F., Tang J., Satz J., Campbell K., Suedhof T.C.;
RT "A stoichiometric complex of neurexins and dystroglycan in brain.";
RL J. Cell Biol. 154:435-445(2001).
CC -1- FUNCTION: NEURONAL CELL SURFACE PROTEIN THAT MAY BE INVOLVED IN
CC CELL RECOGNITION AND CELL ADHESION. MAY MEDIATE INTRACELLULAR
CC SIGNALING.
CC -1- SUBUNIT: LAMININ G-LIKE DOMAIN 2 BINDS TO NEUREXOPHILIN 1.
CC ISOFORMS ALPHA 2C BIND TO ALPHA-DYSTROGLYCAN.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 216 ISOFORMS; MAY BE PRODUCED BY
CC SPLICED EXTRACELLULAR DOMAINS AT SITES 1 TO 5, EACH CONSISTING OF
CC MODULAR SEQUENCES (A-C) THAT SEEM TO BE USED INDEPENDENTLY.
CC ADDITIONAL ISOFORMS MAY DERIVE FROM A MINOR CYTOPLASMIC SPLICE
CC SITE 6. BETA ISOFORMS (AC Q63376) SHARE THE COMBINATION OF
CC ALTERNATIVE SPLICED DOMAINS AT SITES 4, 5 AND 6. SHOWN IS ISOFORM
CC ALPHA 1A2A3A4A5A. ALPHA AND BETA ISOFORMS DIFFER IN THEIR N-
CC TERMINUS BY USE OF ALTERNATIVE PROMOTERS IN THE NRXN2 GENE.
CC -1- TISSUE SPECIFICITY: BRAIN (NEURONAL SYNAPSE).
CC -1- SIMILARITY: CONTAINS 6 LAMININ G-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE NEUREXIN FAMILY.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M96376; AAA41706.1; -.
DR EMBL; M96376; AAA41707.1; -.
DR HSP; Q63373; IC4R.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00008; EGF_3.
DR Pfam; PF00054; laminin_G; 5.
DR SMART; SM00181; EGF_3.
DR SMART; SM00282; LamG; 6.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; FALSE_NEG.
DR PROSITE; PS00025; LAM_G_DOMAIN; 6.
KW Signal; Transmembrane; Repeat; Cell adhesion; Glycoprotein;
KW Alternative splicing; Polymorphism.
FT SIGNAL 1 29 BY SIMILARITY.
FT CHAIN 30 1715 NEUREXIN 2-ALPHA.
FT DOMAIN 30 1639 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1640 1660 POTENTIAL.
FT DOMAIN 1661 1715 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 30 206 LAMININ G-LIKE 1.
FT DOMAIN 202 242 EGF-LIKE 1.
FT DOMAIN 289 486 LAMININ G-LIKE 2.
FT DOMAIN 493 686 LAMININ G-LIKE 3.
FT DOMAIN 690 727 EGF-LIKE 2.
FT DOMAIN 732 907 LAMININ G-LIKE 4.
FT DOMAIN 921 1096 LAMININ G-LIKE 5.
FT DOMAIN 1099 1136 EGF-LIKE 3.
FT DOMAIN 1140 1348 LAMININ G-LIKE 6.
FT DOMAIN 1371 1374 POLY-THR.
FT DOMAIN 1377 1380 POLY-THR.
FT DOMAIN 1447 1450 POLY-PRO.
FT DOMAIN 1647 1650 POLY-ALA.
FT CARBOHYD 60 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 295 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 338 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 844 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPPLIC 250 259 MISSING (IN ISOFORM ALPHA 1B).
FT VARSPPLIC 283 MISSING (IN ISOFORM ALPHA 1C).
FT VARSPPLIC 385 399 MISSING (IN ISOFORM ALPHA 2C).
FT VARSPPLIC 393 399 MISSING (IN ISOFORM ALPHA 2B).
FT VARSPPLIC 797 806 DCLRVGAPS -> G (IN ISOFORM ALPHA 3C).
FT VARSPPLIC 807 809 MISSING (IN ISOFORM ALPHA 3B).
FT VARSPPLIC 1256 1285 MISSING (IN ISOFORM ALPHA 4B).
FT VARSPPLIC 1421 1614 MISSING (IN ISOFORM ALPHA 5B).
FT VARSPPLIC 1667 1715 DEGSQVDQSRNYSISNQSNQNGAVVKKAPAKTPSKAKK
NKDEYYV -> CRKSPREKLLDPSAQGLDLAKACCCV
RCRATCIAGKPLERGGGEGEGEQMQIYIKNK (IN
ISOFORM ALPHA 6).
FT VARIANT 434 434 N -> L.
FT SEQUENCE 1715 AA; 185282 MW; 59FBBF18661F3DB15 CRC64;

Query Match 8.2%; Score 78.5; DB 1; Length 1715;
Best Local Similarity 20.4%; Pred. No. 32;
Matches 37; Conservative 25; Mismatches 60; Indels 59; Gaps 7;

Qy 54 GKYODGSGYEAFTKKEE-----YKDNSRFAICTV-----SLNDALTL-----RAGLAYDK 99
Db 271 GRGAGNVHQTKKEEFVATFRKGNESFCYDLSHNPQSSSTDEITLAFRTLQRGLMLHT 330
Qy 100 AASTHLSASIPD-----TDRMYSIGATYKTPNLSVDV 134
Db 331 GKSANYNLSLKSGAVWLIINLGSAPALVEPVNGKFNDAWHDV-----RVTRNLROHA 386

```



```
Db 217 GNPFVVKDTFLFKNNRNE-----KHLRRKKGH 245
RESULT 30
PORI_BPPA2
ID_ PORI_BPPA2 STANDARD; PRT; 365 AA.
AC P07238;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane porin protein LC precursor.
GN LC.
OS Bacteriophage PA-2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC unclassified Siphoviridae.
OX NCBI_TaxID=10738;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86304457; PubMed=3017988;
RA Blasband A.J., Marcotte W.R. Jr., Schnaitman C.A.;
RT "Structure of the lc and mmpC outer membrane porin protein genes of
RT lambdaoid bacteriophage."
RL J. Biol. Chem. 261:12723-12732(1986).
CC -!- FUNCTION: PORINS ARE MAJOR PROTEINS FOUND IN THE OUTER MEMBRANES
CC OF GRAM-NEGATIVE BACTERIA WHERE THEY FORM CHANNELS FOR THE
CC NONSPECIFIC PERMEATION OF SMALL SOLUTES. (MOLECULES WITH MW LOWER
CC THAN 4000-6000 DALTONS).
CC -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (By similarity).
CC -!- MISCELLANEOUS: A PORIN GENE CAN ALSO BE FOUND IN THE GENOMES OF
CC CERTAIN LAMBDROID BACTERIOPHAGE, AND ITS PROTEIN IS EXPRESSED IN
CC THE LYSGENIC STATE. IN E. COLI THE EXPRESSION OF OMPC AND OMPF
CC PROTEINS IS THEN REDUCED SUBSTANTIALLY.
CC -!- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J02580; AAA32301.1; -
DR PIR; D25647; MMBPP2.
DR HSSP; O52503; IIIV.
DR InterPro; IPR003229; OMP_2.
DR Pfam; IPR001702; Porin_gram-ve.
DR PRINTS; PR00182; ECOLNEIPORIN.
DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 23
FT CHAIN 24 365 OUTER MEMBRANE PORIN PROTEIN LC.
FT CONFLICT 99 99 H -> R (IN REF. 1; AA SEQUENCE).
SQ SEQUENCE 365 AA; 40290 MW; 0FBC0531P9C0205 CRC64;
Query Match 8.08; Score 77; DB 1; Length 365;
Best Local Similarity 20.28; Pred. No. 6.8;
Matches 48; Conservative 27; Mismatches 57; Indels 106; Gaps 10;
Oy 21 AYWELSGFHLTD--QWAIHSYK--YTEMRSFKELRGKYODGSGYEAFTRKEEYKNSR 76
Db 155 ATYRNDFGLVDGLNFAAQYQGNDRSDFDNYTEG--DGFQFSA---TYVEG--- 205
Oy 77 FAIGTYSLND-----ALTRLAGLAYDKRAASKTHLSASIPDTR 115
Db 206 FGIGATYAKSRTDQTQVWAGKVLPEVFASGKNAEVAAGLYD--ANNIYLATTYSETQN 263
Oy 116 M-----WYSIGATY 124
Db 264 MTFADHFVANKQNEAFAVAQYQDFGLRPSVAYLQSKGDKLGVWGDDLVKYVDVGATY 323
```

```
QY 125 KFTPNLSVDVGFAHLRGKKKHFVETQNIKGLLVEADYTK---ATANLYGJNLNRYR 179
DB 324 YFNKNWSIFVDY-----KINLLDKNDFTKALGVSTDDIVAVGLVYQF 365
RESULT 31
FECA_ECOLI
ID FECA_ECOLI STANDARD; PRT; 774 AA.
AC P13036;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Iron(III) dicitrate transport protein fecA precursor.
GN FECA OR B4291.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 34-40.
RC STRAIN=B.
RX MEDLINE=88227855; PubMed=2836368;
RA Pressier U., Staudenmaier H., Zimmermann L., Braun V.;
RT "Genetics of the iron dicitrate transport system of Escherichia
RT coli."
RL J. Bacteriol. 170:2716-2724(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes."
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [3]
RP SEQUENCE OF 1-428 FROM N.A.
RX MEDLINE=91072220; PubMed=2254251;
RA van Hove B., Staudenmaier H., Braun V.;
RT "Novel two-component transmembrane transcription control: regulation
RT of iron dicitrate transport in Escherichia coli K-12."
RL J. Bacteriol. 172:6749-6758(1990).
RN [4]
RP SEQUENCE OF 731-774 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=89213950; PubMed=2651410;
RA Staudenmaier H., van Hove B., Yaraghi Z., Braun V.;
RT "Nucleotide sequences of the fecSCDE genes and locations of the
RT proteins suggest a periplasmic-binding-protein-dependent transport
RT mechanism for iron(III) dicitrate in Escherichia coli."
RL J. Bacteriol. 171:2626-2633(1989).
CC -!- FUNCTION: FECA IS THE OUTER MEMBRANE RECEPTOR PROTEIN IN THE
CC IRON(III) DICITRATE TRANSPORT SYSTEM.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- INDUCTION: FOR INDUCTION THE TONB AND THE EXBB PROTEIN HAVE TO
CC BE ACTIVE. REGULATION BY THE IRON LEVEL MEDIATED BY THE FUR PROTEIN
CC AND INDUCTION BY CITRATE PLUS IRON SUGGEST THAT THE IRON(III)
CC DICITRATE COMPLEX MUST ENTER THE PERIPLASM WHERE IT BINDS TO A
CC TRANSMEMBRANE PROTEIN, WHICH REGULATES DIRECTLY OR VIA A FURTHER
CC INDUCTOR, TRANSCRIPTION OF THE FEC GENES.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M20981; AAA23760.1; -
DR EMBL; U14003; AAA97187.1; -
```

```
DR EMBL: A5000499; AAC77247.1; -.
DR EMBL: M63115; AAA23768.1; -.
DR EMBL: M26397; AAA23761.1; -.
DR PIR: JY0022; ORECEA.
DR EcoGene: EGI0286; fecA.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC; 1.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW Complete proteome.
FT SIGNAL 1 33
FT CHAIN 34 774
FT IIRON(III) DICITRATE TRANSPORT PROTEIN
FECA.
FT SITE 56 63
FT SITE 757 774 TONB C-TERMINAL BOX.
FT CONFLICT 16 16 T -> A (IN REF. 1).
FT CONFLICT 190 190 A -> T (IN REF. 1).
FT CONFLICT 300 300 T -> M (IN REF. 1 AND 3).
FT CONFLICT 357 357 W -> R (IN REF. 1).
FT CONFLICT 444 444 L -> V (IN REF. 1).
FT CONFLICT 749 749 R -> A (IN REF. 1).
SQ SEQUENCE 774 AA; 85321 MW; 16B5B510276C3B09 CRC64;

Query Match 8.0%; Score 76.5; DB 1; Length 774;
Best Local Similarity 20.7%; Pred. No. 19;
Matches 51; Conservative 32; Mismatches 68; Indels 95; Gaps 12;

QY 1 QHNGVLGPGYIGKSLTLKLPAYWELSGFHOLTQDWAHYHYKKTWS----- 47
DB 508 QNNAITGTH-EEVSYNAPLPA---LNVLYHLTDSWNL---YANTEGSFGTVQVYSQIGKAV 560
QY 48 -----RFKELRGKYQDGS-----GYEAFETKKEEYKDN----- 74
DB 561 QSCNVEPEKARTWELGTRYDDGALTAEMGLFLINFNQYDSNOTNDRVTARGKTRHTGLE 620
QY 75 --SRFAIGTYSNDALTTLRAGLAY-----DKAASKTHLSASIPDPDRMYSIGATYK- 125
DB 621 TQARYDLGTLTPTLDNYSIYAYVNAEIREKRGDTYGNL---VPFSPKHGTLGVQYK 677
QY 126 --FTPNLSVDVGFALHGRGKKHFEVTONIKGLLLVEADYTKATAN-----LYGL 173
DB 678 GNTFNLSNDF-----QSSQFADNAN-----TVKESADGSTGRIPGFMLWGA 719
QY 174 NLNRYE 179
DB 720 RVAYDF 725

RESULT 32
POTI_RHOBL
ID POTI_RHOBL STANDARD; PRT; 289 AA.
AC P39767;
DT 01-FEB-1995 (Rel. 31, Created).
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Porin.
DE OPMA.
OS Rhodopsedomonas blastica.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1075;
RN [1]
RX MEDLINE=94191532; PubMed=8142898;
RA Kreusch A., Neubueser A., Schiltz E., Weckesser J., Schulz G.E.;
RT "Structure of the membrane channel porin from Rhodopsedomonas
RL blastica at 2.0-A resolution.";
RL Protein Sci. 3:58-63(1994).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.96 ANGSTROMS).
RX MEDLINE=95055730; PubMed=7525973;
RA Kreusch A., Schulz G.E.;
```

```
RT "Refined structure of the porin from Rhodopsedomonas blastica.
RT Comparison with the porin from Rhodobacter capsulatus.";
RL J. Mol. Biol. 243:891-905(1994).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF MUTANTS.
RX MEDLINE=98348035; PubMed=9684893;
RA Schmid B., Maveyraud L., Kromer M., Schulz G.E.;
RT "Porin mutants with new channel properties.";
RL Protein Sci. 7:1603-1611(1998).
CC -!- FUNCTION: FORMS CHANNELS THAT ALLOW THE PASSIVE DIFFUSION OF SMALL
CC HYDROPHILIC SOLUTES UP TO AN EXCLUSION LIMIT OF ABOUT 0.6 kDa.
CC -!- SUBUNIT: HOMOTRIMER.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
DR PIR: S38806; S38806.
DR PDB: 1PRN; 24-OCT-94.
DR PDB: 2PRN; 13-JAN-99.
DR PDB: 3PRN; 12-AUG-98.
DR PDB: 5PRN; 12-AUG-98.
DR PDB: 6PRN; 12-AUG-98.
DR PDB: 7PRN; 12-AUG-98.
DR PDB: 8PRN; 12-AUG-98.
DR PDB: 1BH3; 12-AUG-98.
DR InterPro: IPR001702; Porin_gram-ve.
DR Pfam: PF00267; Gram-ve_porins; 1.
KW Outer membrane; Transmembrane; Porin; 3D-structure.
SQ SEQUENCE 289 AA; 30597 MW; 08252D9803A1044C CRC64;

Query Match 7.9%; Score 76; DB 1; Length 289;
Best Local Similarity 32.2%; Pred. No. 63;
Matches 29; Conservative 7; Mismatches 28; Indels 26; Gaps 5;

QY 73 DNSRFAIGTYSNDALTTLRAGLA-YDKAAS-----KTHLSASIPDTRM 116
DB 190 DNDIAFGAAYKENDACTV--GLNWDYDGLSTAGDQVTLGNYAFGATTVPAYVSDIDRA 247
QY 117 W-----YSIGATYKTPNL-----SVDVGFH 138
DB 248 GADTAYGIGADYQFAEGVKVSGVQSGFAN 277

RESULT 33
ARHY_HUMAN
ID ARHY_HUMAN STANDARD; PRT; 357 AA.
AC P54922;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADP-ribosylarginine hydrolase (EC 3.2.2.19) (ADP-ribose-L-arginine
DE cleaving enzyme).
GN ADPRH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=93352593; PubMed=8349667;
RA Takada T., Iida K., Moss J.;
RT "Cloning and site-directed mutagenesis of human ADP-ribosylarginine
RT hydrolase.";
RL J. Biol. Chem. 268:17837-17843(1993).
CC -!- FUNCTION: IT CATALYZES REVERSE REACTION OF MONO-ADP-RIBOSYLATION.
CC -!- CATALYTIC ACTIVITY: N(2)-(ADP-D-ribosyl)-L-arginine + H(2)O -> L-
CC arginine + ADP-ribose.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```



```

Db 292 LRP5LAYLQSGK-NLGRGYDDEDLTKYVDVGATYFNKNMSTYVDY----- 337
Qy 150 ONIKGLLLVEADYTKA---TANLYGLNLNRYF 179
Db 338 ---KINLLDNOFTRDAGINTDNIVALGLVYQF 367

RESULT 35
OMPC_SALTY
ID OMPC_SALTY STANDARD; PRT; 378 AA.
AC 052503; P09878;
DT 01-MAR-1989 (Rel. 10, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Outer membrane protein C precursor (Porin ompC).
GN OMPC OR SPM2267 OR STY2493.
OS Salmonella typhimurium, and
OC Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=ATCC 14028;
RX MEDLINE=99457705; PubMed=10528398;
RA Negm R.S., Pistole T.G.;
RT "The porin OmpC of Salmonella typhimurium mediates adherence to
macrophages";
RL Can. J. Microbiol. 45:658-669(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi;
RX MEDLINE=90060831; PubMed=2684785;
RA Puente J.L., Alvarez-Scherer V., Gosset G., Calva E.;
RT "Comparative analysis of the Salmonella typhi and Escherichia coli
ompC genes.";
RL Gene 83:197-206(1989).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=Ty2;
RX MEDLINE=8831959; PubMed=3412902;
RA Venegas A., Gomez I., Zaror I., Yudelevich A.;
RT "The nucleotide sequence of the Salmonella typhi ompC porin gene.";
RL Nucleic Acids Res. 16:7721-7721(1988).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,
Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [6]

```

```

RP 3D-STRUCTURE MODELING, AND CRYSTALLIZATION.
RC SPECIES=S.typhi;
RX MEDLINE=99332091; PubMed=10405180;
RA Arockiasamy A., Krishnaswamy S.;
RT "Crystallization of the immunodominant outer membrane protein OmpC;
the first protein crystals from Salmonella typhi, a human pathogen.";
RL FEBS Lett. 453:380-382(1999).
CC -|- FUNCTION: FORMS PASSIVE DIFFUSION PORES WHICH ALLOW SMALL
MOLECULAR WEIGHT HYDROPHILIC MATERIALS ACROSS THE OUTER MEMBRANE.
CC -|- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -|- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF039309; AAB96675.1; -
DR EMBL: AE008801; AAL21169.1; -
DR EMBL: M31424; AAA27169.1; -
DR EMBL: X07835; CAA30688.1; -
DR EMBL: AL627274; CAD07499.1; -
DR PIR: JQ0119; MMEBPC.
DR PIR: S01248; S01248.
DR PDB: 1IIV, 14-OCT-98.
DR StyGene: SG10675; ompC.
DR InterPro: IPR003229; OMP_2.
DR InterPro: IPR001702; Porin_gram-ve.
DR Pfam: PF00267; Gram-ve-porins; 1.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Phage recognition; Signal;
KW 3D-structure; Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 378 OUTER MEMBRANE PROTEIN C.
FT CONFLICT 42 42 H -> N (IN REF. 1).
FT CONFLICT 223 223 N -> D (IN REF. 1).
FT CONFLICT 253 253 L -> V (IN REF. 1).
FT CONFLICT 258 258 S -> F (IN REF. 1).
FT CONFLICT 340 340 K -> T (IN REF. 1).
FT CONFLICT 356 357 ND -> YE (IN REF. 1).
FT CONFLICT 362 362 MISSING (IN REF. 4).
SQ SEQUENCE 378 AA; 41239 MW; 58479E8685B43ED1 CRC64;

Query Match 7.9%; Score 75.5; DB 1; Length 378;
Best Local Similarity 27.2%; Pred. No. 9.7;
Matches 41; Conservative 17; Mismatches 60; Indels 33; Gaps 8;

Qy 36 ATHYSKYTEWSREKELRGKYODGSGYEATFKKEEYKONSREAGTGYSLNDALTLRAGL 95
Db 254 AAQYSQTYNA-TRFGTSNG--SNPSTSYGFANKAQ-----NFEVVAQYQFD---FGLRPSV 303
Qy 96 AYDKAASK---THLSASIPDTRMWY-SIGATYKFTPNLSVDVYGFALRGKKHFEVTON 151
Db 304 AVLQSKGKDINSYGASGYGDQDIVKYVDVGATYFYNKNMSTYVDY----- 348
Qy 152 IKGLLLVEADYTKA---TANLYGLNLNRYF 179
Db 349 -KINLLDKNDFTRDAGINTDDIVALGLVYQF 378

RESULT 36
OPAG_NEIGO
ID OPAG_NEIGO STANDARD; PRT; 247 AA.
AC Q04875;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Opacity protein OPA52 precursor (OPA31) (Fragment).
GN OPAG.

```





GN VASI OR YGR094W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 NCBI\_TaxID=4932;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=87222321; PubMed=3294828;  
 RX Jordana X., Chatton B., Paz-Weisshaar M., Buhler J.-M., Cramer F.,  
 RA Ebel J.-P., Fasiolo F.;  
 RA "Structure of the yeast valyl-tRNA synthetase gene (VASI) and the  
 RT homology of its translated amino acid sequence with Escherichia coli  
 RT isoleucyl-tRNA synthetase.";  
 RL J. Biol. Chem. 262:7189-7194 (1987).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP Wedler H., Scharfe M., Wedler E., Wambutt R.;  
 RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RP Hernandez K., Weber N., Wipfli P., Schmidheini T.;  
 RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RN SEQUENCE OF 1-78 FROM N.A.  
 RP MEDLINE=88087140; PubMed=3275649;  
 RX Chatton B., Walter P., Ebel J.-P., Lacroite F., Fasiolo F.;  
 RA "The yeast VASI gene encodes both mitochondrial and cytoplasmic  
 RT valyl-tRNA synthetases.";  
 RL J. Biol. Chem. 263:53-57 (1988).  
 CC -|- CATALYTIC ACTIVITY: ATP + L-valine + tRNA(Val) = AMP + diphosphate  
 CC + L-valyl-tRNA(Val).  
 CC -|- SUBCELLULAR LOCATION: Mitochondrial and cytoplasmic.  
 CC -|- ALTERNATIVE PRODUCTS: 2 isoforms; a mitochondrial form (shown  
 CC here) and a cytoplasmic form; are produced by alternative  
 CC initiation.  
 CC -|- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; J02719; AAA35207.1; -  
 DR EMBL; J2879; CAA97097.1; -  
 DR EMBL; M18392; AAA35205.1; -  
 DR EMBL; M18392; AAA35206.1; -  
 DR PIR; A29871; SYBYT.  
 DR HSP; P96142; IGAX.  
 DR SGD; S0003326; VAS1.  
 DR InterPro; IPR002300; tRNA-synt\_la.  
 DR InterPro; IPR001412; tRNA-synt\_I.  
 DR InterPro; IPR002303; tRNA-synt\_val.  
 DR Pfam; PF00133; tRNA-synt\_1; 1.  
 DR PRINTS; P00986; TRNASYNTHAL.  
 DR TIGRFAMs; TIGR00422; vals; 1.  
 DR PROSITE; PS00178; AA.TRNA.LIGASE.I; 1.  
 DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KW Mitochondrion; Transit peptide; Alternation initiation.  
 FT TRANSIT 1 47  
 FT CHAIN 48 1104  
 FT CHAIN 47 1104  
 FT VALYL-TRNA SYNTHETASE, MITOCHONDRIAL  
 FT ISOFORM.  
 FT VALYL-TRNA SYNTHETASE, CYTOPLASMIC  
 FT ISOFORM.  
 FT INIT\_MET 47  
 FT SITE 190 200  
 FT "HIGH" REGION.  
 FT BINDING 703 707  
 FT "KMSKS" REGION.  
 FT SITE 706 706  
 FT BINDING 147 147  
 FT CONFLICT 147 147  
 FT CONFLICT 540 540  
 FT CONFLICT 540 540  
 FT R -> K (IN REF. 1).  
 FT R -> K (IN REF. 1).  
 FT SEQUENCE 1104 AA; 125769 MW; 6493AEF37ECD4A7C CRC64;

Query Match 7.8%; Score 75; DB 1; Length 1104;  
 Best Local Similarity 20.7%; Pred. No. 40;  
 Matches 36; Conservative 25; Mismatches 63; Indels 50; Gaps 6;  
 QY 20 PAYWELSGFHOLDTDMWAIHYSKYKTEMSRFEKLRGYQDGSYGAEFTKKEEYKDNRS--- 76  
 DB 812 PATEGLSGNESSLVKWLH---KLTETSKVN-----EALDKRDFLTSTSSYE 857  
 QY 77 --FAITTTSLNDALTLRAGLAYDKAASKTHLSASIPD-----TDRMWSIGA 122  
 DB 858 FWYLICDVVIENSKYLIQEGSAIEKKSADTLVLLDNALKLIHPPMPFISEEMWQRL-- 915  
 QY 123 TYKFTPNLSVDVGFALHKKHVFETQNIKGLLLVADYTTKATKATLYGLNLN 176  
 DB 916 -----PKRSTKASIVKASYPYV-----SEYDDVKSNAYDLVLN 952  
 RESULT 39  
 POLG.PEMVM  
 ID POLG.PEMVM STANDARD; PRT; 3099 AA.  
 AC O56075;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Genome polyprotein [Contains: N-terminal protein (P1); Helper  
 DE component proteinase (EC 3.4.22.45) (HC-Pro); Protein P3; 6 kDa  
 DE protein 1 (6K1); Cytoplasmic inclusion protein (CI); 6 kDa protein 2  
 DE (6K2); Genome-linked protein (VPG); Nuclear inclusion protein A (NI-A)  
 DE (NIA) (EC 3.4.22.44) (49 kDa proteinase) (49 kDa-Pro); Nuclear  
 DE inclusion protein B (NI-B) (NIB) (RNA-directed RNA polymerase) (RC  
 DE 2.7.7.48); Coat protein (CP)].  
 OS Peanut mottle virus (strain M).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;  
 OC Potyvirus.  
 OC NCBI\_TaxID=103926;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RA Flasinaki S., Gonzales R.A., Cassidy B.G.;  
 RT "The complete nucleotide sequence of peanut mottle virus (M strain)  
 RT genomic RNA.";  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 CC -|- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID  
 CC TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.  
 CC -|- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT  
 CC MAY BE INVOLVED IN REPLICATION.  
 CC -|- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.  
 CC -|- CATALYTIC ACTIVITY: Hydrolyzes glutamyl bonds, and activity is  
 CC further restricted by preferences for the amino acids in P6 - P1'  
 CC that vary with the species of potyvirus, e.g. Glu-Xaa-Xaa-Tyr-Xaa-  
 CC Glu+(Ser or Gly) for the enzyme from tobacco etch virus. The  
 CC natural substrate is the viral polyprotein, but other proteins and  
 CC oligopeptides containing the appropriate consensus sequence are  
 CC also cleaved.  
 CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC [RNA](N).  
 CC -|- CATALYTIC ACTIVITY: Hydrolyzes a Gly-I-Gly bond at its own C-  
 CC terminus, commonly in the sequence -Tyr-Xaa-Val-Gly-I-Gly, in the  
 CC processing of the potyviral polyprotein.  
 CC -|- PTM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.  
 CC -|- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE  
 CC POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC  
 CC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT  
 CC INDIVIDUAL PROTEINS.  
 CC -|- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.  
 CC -|- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.  
 CC -|- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AF023848; AAB94595.1; -

DR MEROPS; S30.001; -

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR001730; Peptidase\_C4.

DR InterPro; IPR001456; Peptidase\_C6.

DR InterPro; IPR002540; Poty\_P1.

DR InterPro; IPR001592; Poty\_coat.

DR InterPro; IPR001205; RNA\_pol\_P3D.

DR InterPro; IPR001254; Ser\_protease\_Try.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.

DR Pfam; PF00767; Poty\_coat; 1.

DR Pfam; PF00851; Peptidase\_C6; 1.

DR Pfam; PF00863; Peptidase\_C4; 1.

DR Pfam; PF01577; Poty\_P1; 1.

DR PRINTS; PR00966; NIAPOTYPTASE.

DR SMART; SM00487; DEXdc; 1.

DR SMART; SM00490; HELIC; 1.

KW Hydrolase; Transferase; Thiol protease; RNA-directed RNA polymerase;

KW Coat protein; Polyprotein; Covalent protein-RNA linkage; Helicase;

KW ATP-binding.

FT CHAIN 1 322 N-TERMINAL PROTEIN.

FT CHAIN 323 779 HELPER COMPONENT PROTEINASE.

FT CHAIN 780 1128 PROTEIN P3.

FT CHAIN 1129 1180 6 KDA PROTEIN 1.

FT CHAIN 1181 1814 CYTOPLASMIC INCLUSION PROTEIN.

FT CHAIN 1815 1867 6 KDA PROTEIN 2.

FT CHAIN 1868 ? GENOME-LINKED PROTEIN.

FT CHAIN ? 2303 NUCLEAR INCLUSION PROTEIN A.

FT CHAIN 2304 2821 NUCLEAR INCLUSION PROTEIN B.

FT CHAIN 2822 3099 COAT PROTEIN.

FT CHAIN 1868 2303 PUTATIVE NUCLEAR INCLUSION PROTEIN A.

FT NP\_BIND 1265 1272 ATP (POTENTIAL).

FT NP\_SEQUENCE 3099 AA; 351032 MW; OD8E9FC7603FOA4B CRC64;

Query Match 7.88; Score 75; DB 1; Length 3099;

Best Local Similarity 28.78; Pred. No. 1.4e+02;

Matches 33; Conservative 11; Mismatches 33; Indels 38; Gaps 6;

Qy 49 FKELRGKYDGS-GYEATKKEEYKDSRFAIG-----TYSINDALTLRAG 94

Db 2355 FKPLMGAYQPSKLNKEAFT-KDLKYQNEIIVGEVDKDFNAVEAVIYLLD----- 2406

Qy 95 LAYDKAASKTHLSASIPDTRMWSYIGATYKFTPNLSVDVGFALHGRKKHFVET 149

Db 2407 LGFGECAYVTDEEA-ILDSLNKKAAGALYK-----GKKKEYFES 2445

RESULT 40

OPAA\_NEIGO STANDARD; PRT; 239 AA.

AC Q04876;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Opacity protein OPA53 precursor (Fragment).

GN OPAA.

OS Neisseria gonorrhoeae.

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

OX NCBI\_TaxID=485;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MS11 / F3;

RA MEDLINE=93178439; PubMed=8440254;

RX Kupsch E.-M., Knepper B., Kuroki T., Heuer I., Meyer T.F.;

RT "Variable opacity (Opa) outer membrane proteins account for the cell

RT trophisms displayed by Neisseria gonorrhoeae for human leukocytes and

RT epithelial cells."

RL EMBO J. 12:641-650(1993).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=MS11 / V18;

RX MEDLINE=92114767; PubMed=1815562;

RA Bhat K.S., Gibbs C.P., Barrera O., Morrison S.G., Jaehnig F.,

RA Stern A., Kupsch E.-M., Meyer T.F., Swanson J.;

RT "The opacity proteins of Neisseria gonorrhoeae strain MS11 are

RL encoded by a family of 11 complete genes."

RL Mol. Microbiol. 5:1889-1901(1991).

RN [3]

RP ERRATUM.

RX MEDLINE=92261323; PubMed=1584024;

RA Bhat K.S., Gibbs C.P., Barrera O., Morrison S.G., Jaehnig F.,

RA Stern S., Kupsch E.-M., Meyer T.F., Swanson J.;

RL Mol. Microbiol. 6:1073-1076(1992).

CC -1- FUNCTION: IMPLICATED IN A NUMBER OF ADHERENCE FUNCTIONS. OPA

CC PROTEINS ARE IMPLICATED IN PATHOGENESIS AND ARE SUBJECT TO PHASE

CC VARIATION.

CC -1- SUBCELLULAR LOCATION: Outer membrane.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; Z18930; CAA79363.1; -

CC EMBL; X60709; CAA43119.1; -

DR PIN; S28630; S28630.

DR PIN; S36331; S36331.

DR InterPro; IPR003394; Porin\_opacity.

DR Pfam; PF02462; Opacity; 1.

KW Outer membrane; Multigene family; Signal.

FT NON\_TER 1 1

FT SIGNAL <1 1 POTENTIAL.

FT CHAIN 2 >239 OPACITY PROTEIN OPA53.

FT VARIANT 92 92 N -> D (IN MS11 / V18).

FT NON\_TER 239 239

SQ SEQUENCE 239 AA; 26734 MW; 2F12B69DEB26CAA0 CRC64;

Query Match 7.88; Score 74.5; DB 1; Length 239;

Best Local Similarity 24.08; Pred. No. 6.9;

Matches 35; Conservative 17; Mismatches 41; Indels 53; Gaps 8;

Qy 35 WAIHVS-KYTEWSR-----FKELRGKYDQSGYEAF-----TKKEEYKDSRFAIGTY 83

Db 65 WRIAADYARYRKWNNSKYSYSIKEL-GRNDNSTNSSHLNKTOKTQTHOENGTF----- 117

Qy 84 SLNDALTLRAGLAYDKAASKTHLSASIPDTRMWSYIGATYKFTPNLSVDVGFALHGRKK 143

Db 118 -----HATSSLSLSAI-----YDFKLNKFKPYICVRVAYGHV----- 150

Qy 144 KHFVETONIKGLLLVEADYTTKATAN 169

Db 151 KHQVRS-----VESETTTVTTHN 168

Search completed: May 12, 2003, 09:40:40

Job time : 19 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 09:38:54 ; Search time 30 Seconds.  
(without alignments)  
1229.414 Million cell updates/sec

Title: US-09-995-493-52  
Perfect score: 957  
Sequence: 1 OHNGVLGPYIGKSGSLTLP.....ADYTKATANLYGLNLNVR 179

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*

- 1: sp-archaea.\*
- 2: sp-bacteria.\*
- 3: sp-fungi.\*
- 4: sp-human.\*
- 5: sp-invertebrate.\*
- 6: sp-mammal.\*
- 7: sp-mhc.\*
- 8: sp-organelle.\*
- 9: sp-phage.\*
- 10: sp-plant.\*
- 11: sp-rodent.\*
- 12: sp-virus.\*
- 13: sp-vertebrate.\*
- 14: sp-unclassified.\*
- 15: sp-rvirus.\*
- 16: sp-bacteriap.\*
- 17: sp-archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	580.5	60.7	427	Q48051	Q48051 haemophilus
2	571.5	59.7	449	Q9K2Q4	Q9K2Q4 haemophilus
3	571.5	59.7	449	Q9KHF9	Q9KHF9 haemophilus
4	570.5	59.6	449	Q00754	Q00754 haemophilus
5	564	58.9	451	Q9KHF8	Q9KHF8 haemophilus
6	563.5	58.9	451	Q9KHF7	Q9KHF7 haemophilus
7	562.5	58.8	449	Q9KHF5	Q9KHF5 haemophilus
8	561	58.6	455	Q9KHF2	Q9KHF2 haemophilus
9	561	58.6	455	Q9KHH1	Q9KHH1 haemophilus
10	559.5	58.5	456	Q9KHG3	Q9KHG3 haemophilus
11	559	58.4	455	Q9KHH4	Q9KHH4 haemophilus
12	559	58.4	455	Q9KHH3	Q9KHH3 haemophilus
13	559	58.4	455	Q00450	Q00450 haemophilus
14	557	58.2	455	Q9KHH5	Q9KHH5 haemophilus
15	553.5	57.8	451	Q9KHF6	Q9KHF6 haemophilus
16	552.5	57.7	450	Q9KHF4	Q9KHF4 haemophilus

17	552.5	57.7	450	2	Q9KHF3	Q9KHF3 haemophilus
18	550.5	57.5	451	2	Q9KHG0	Q9KHG0 haemophilus
19	549.5	57.4	456	2	Q9KHG9	Q9KHG9 haemophilus
20	549	57.4	456	2	Q9KHH6	Q9KHH6 haemophilus
21	549	57.4	456	2	Q9KHG7	Q9KHG7 haemophilus
22	547	57.2	451	2	Q9KHG2	Q9KHG2 haemophilus
23	547	57.2	459	2	Q9K324	Q9K324 haemophilus
24	547	57.2	459	2	Q9KHH8	Q9KHH8 haemophilus
25	547	57.2	459	2	Q9KHU7	Q9KHU7 haemophilus
26	545.5	57.0	456	2	Q9KHG8	Q9KHG8 haemophilus
27	544	56.8	456	2	Q9KHI4	Q9KHI4 haemophilus
28	544	56.8	456	2	Q9KHI2	Q9KHI2 haemophilus
29	542.5	56.7	457	2	Q9KHG6	Q9KHG6 haemophilus
30	538	56.2	456	2	Q9KHH9	Q9KHH9 haemophilus
31	538	56.2	458	2	Q9K397	Q9K397 haemophilus
32	534	55.8	456	2	Q9KHI3	Q9KHI3 haemophilus
33	533.5	55.7	456	2	Q9KHG5	Q9KHG5 haemophilus
34	533.5	55.7	456	2	Q9KHG4	Q9KHG4 haemophilus
35	533	55.7	454	2	Q9KHG1	Q9KHG1 haemophilus
36	533	55.7	456	2	Q9KHI1	Q9KHI1 haemophilus
37	532	55.6	460	2	Q9KHH0	Q9KHH0 haemophilus
38	517	54.0	456	2	Q9KHI0	Q9KHI0 haemophilus
39	421	44.0	412	16	Q9KL84	Q9KL84 vibrio chol
40	372.5	38.9	423	16	Q8ZD48	Q8ZD48 yersinia pe
41	361.5	37.8	442	2	Q9FOY5	Q9FOY5 enterobacte
42	338	35.3	437	16	Q8ZNA5	Q8ZNA5 salmonella
43	338	35.3	437	16	Q8ZAY8	Q8ZAY8 salmonella
44	334	34.9	432	16	Q9KT63	Q9KT63 vibrio chol
45	332	34.7	448	16	Q8XCN6	Q8XCN6 escherichia

ALIGNMENTS

RESULT 1

Q48051	PRELIMINARY;	PRT;	427 AA.
ID	Q48051		
AC	Q48051;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-JUN-2000 (TrEMBLrel. 14, Last annotation update)		
DE	Outer membrane protein P1 (Fragment).		
OS	Haemophilus influenzae.		
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;		
OC	Haemophilus.		
OX	NCBI_TaxID=727;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=F3029;		
RX	MEDLINE=96355877; PubMed=8751915;		
RA	Reed R.B., Frost J.B., Kort K., Myers S.D., Lesse A.J.;		
RT	"DNA sequence analysis and restriction fragment length polymorphisms of the P1 gene of Haemophilus influenzae biogroup aegyptius associated with Brazilian purpuric fever.";		
RT	Infect. Immun. 64:3666-3672(1996).		
DR	EMBL; U28808; AAC44459.1; -		
FT	NON-TER	1	
FT	VARIANT	61	61
FT	VARIANT	265	265
FT	VARIANT	385	385
SQ	SEQUENCE	427 AA;	80620B4A6FA0DD27 CRC64;

Query Match	60.7%;	Score	580.5;	DB	2;	Length	427;
Best Local Similarity	63.7%;	Pred. No.	1.3e-42;				
Matches	107;	Conservative	31;	Mismatches	27;	Indels	3;
						Gaps	2;
QY	12	KGSLTKLPAYWELSGHQLTDQWAIHYKYKTEWSRFEKLRGKYQDGSYEAFTKKEEY	71				
Db	263	KGDLTKLPDYLELSGPHQLTDKFAVHSYKYTHWSRLTKLHASFEDGK--KAFDKELQY	320				
QY	72	KDNSRFAIGTYTYSUNDALTURAGLAYDKAASKTHLSASIPDTRMWSYIGATYKFTPNLS	131				
Db	321	SNNSRVALGASYNLDEKLTLRAGIAMDQAASRRHRSAAIPDTRTWYSLGATYKFTPNLS	380				





```
Best Local Similarity 62.28; Pred. No. 7.2e-41;
Matches 107; Conservative 30; Mismatches 29; Indels 6; Gaps 3;

Qy 12 KGSLLKLPAYWELSGFHQLTDQWAIHYSYKYTWMSRFEKELRGKYQDGSYEATKKEEY 71
Db 286 KGNLTLPDYLELSGFHQLTDKFAVHYSYKYTHWSRLTKLHASFEDGK--KAFDKELQY 343
Qy 72 KDNSRFAIGTTSYSLNDALTLRAGLAYDKAASKTHLSASIPDTRMWYSIGATYKFTPNLS 131
Db 344 SNNRVALGASYNLDEKLTLAGIAYDQAAASRRHRSRAIPDTRTWISLGLATYKFTPNLS 403
Qy 132 VDVGFAHLRGKKHVFETQNI---KGLLL-VEADYTTKATANLYGLNLYRF 179
Db 404 VDLGAYLKGKKVHFKVEKVTIGDKRALTLNTTANYTSQAHANLYGLNLYSF 455

RESULT 9
Q9KHH1 ID Q9KHH1 PRELIMINARY; PRT; 455 AA.
AC Q9KHH1;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Trung-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260354; AAF97568.1; -.
SQ SEQUENCE 455 AA; 49123 MW; FD5882979F10EB24 CRC64;

Query Match 58.68; Score 561; DB 2; Length 455;
Best Local Similarity 62.28; Pred. No. 7.2e-41;
Matches 107; Conservative 30; Mismatches 29; Indels 6; Gaps 3;

Qy 12 KGSLLKLPAYWELSGFHQLTDQWAIHYSYKYTWMSRFEKELRGKYQDGSYEATKKEEY 71
Db 286 KGNLTLPDYLELSGFHQLTDKFAVHYSYKYTHWSRLTKLHASFEDGK--KAFDKELQY 343
Qy 72 KDNSRFAIGTTSYSLNDALTLRAGLAYDKAASKTHLSASIPDTRMWYSIGATYKFTPNLS 131
Db 344 SNNRVALGASYNLDEKLTLAGIAYDQAAASRRHRSRAIPDTRTWISLGLATYKFTPNLS 403
Qy 132 VDVGFAHLRGKKHVFETQNI---KGLLL-VEADYTTKATANLYGLNLYRF 179
Db 404 VDLGAYLKGKKVHFKVEKVTIGDKRALTLNTTANYTSQAHANLYGLNLYSF 455

RESULT 10
Q9KHG3 ID Q9KHG3 PRELIMINARY; PRT; 456 AA.
AC Q9KHG3;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Trung-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260354; AAF97568.1; -.
SQ SEQUENCE 456 AA; 49133 MW; OCB00B162A6A6CB1 CRC64;

Query Match 58.48; Score 559; DB 2; Length 455;
Best Local Similarity 61.68; Pred. No. 1.1e-40;
Matches 106; Conservative 29; Mismatches 31; Indels 6; Gaps 2;

Qy 12 KGSLLKLPAYWELSGFHQLTDQWAIHYSYKYTWMSRFEKELRGKYQDGSYEATKKEEY 71
Db 286 KGNLTLPDYLELSGFHQLTDKFAVHYSYKYTHWSRLTKLHASFEDGK--KAFDKELQY 343
Qy 72 KDNSRFAIGTTSYSLNDALTLRAGLAYDKAASKTHLSASIPDTRMWYSIGATYKFTPNLS 131
Db 344 SNNRVALGASYNLDEKLTLAGIAYDQAAASRRHRSRAIPDTRTWISLGLATYKFTPNLS 403
Qy 132 VDVGFAHLRGKKHVFETQNIKG---LLLVADYTTKATANLYGLNLYRF 179
Db 404 VDLGAYLKGKKVHFKVEKVTIGDKRTLTNTTANYTSQAHANLYGLNLYSF 455

RESULT 12
```

```
SEQUENCE FROM N.A.
RC STRAIN=1-H-1085;
RX MEDLINE=20359342; Pubmed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Trung-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260354; AAF97578.1; -.
SQ SEQUENCE 456 AA; 49133 MW; OD6AFC5DFEB35014 CRC64;

Query Match 58.58; Score 559.5; DB 2; Length 456;
Best Local Similarity 62.08; Pred. No. 9.8e-41;
Matches 106; Conservative 28; Mismatches 32; Indels 5; Gaps 2;

Qy 12 KGSLLKLPAYWELSGFHQLTDQWAIHYSYKYTWMSRFEKELRGKYQDGSYEATKKEEY 71
Db 288 KGNLTLPDYLELSGFHQLTDKFAVHYSYKYTHWSRLTKLYASYENGK--KAFDKELQY 345
Qy 72 KDNSRFAIGTTSYSLNDALTLRAGLAYDKAASKTHLSASIPDTRMWYSIGATYKFTPNLS 131
Db 346 SNNRVALGASYNLDEKLTLAGIAYDQAAASRRHRSRAIPDTRTWISLGLATYKFTPNLS 405
Qy 132 VDVGFAHLRGKKHVFETQNIKGLLL---VEADYTTKATANLYGLNLYRF 179
Db 406 VDLGAYLKGKKVHFKVEKVTISQAQLTLNTTANYTSQAHANLYGLNLYSF 456

RESULT 11
Q9KHH4 ID Q9KHH4 PRELIMINARY; PRT; 455 AA.
AC Q9KHH4;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Trung-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260349; AAF97563.1; -.
SQ SEQUENCE 455 AA; 49111 MW; OCB00B162A6A6CB1 CRC64;

Query Match 58.48; Score 559; DB 2; Length 455;
Best Local Similarity 61.68; Pred. No. 1.1e-40;
Matches 106; Conservative 29; Mismatches 31; Indels 6; Gaps 2;

Qy 12 KGSLLKLPAYWELSGFHQLTDQWAIHYSYKYTWMSRFEKELRGKYQDGSYEATKKEEY 71
Db 286 KGNLTLPDYLELSGFHQLTDKFAVHYSYKYTHWSRLTKLHASFEDGK--KAFDKELQY 343
Qy 72 KDNSRFAIGTTSYSLNDALTLRAGLAYDKAASKTHLSASIPDTRMWYSIGATYKFTPNLS 131
Db 344 SNNRVALGASYNLDEKLTLAGIAYDQAAASRRHRSRAIPDTRTWISLGLATYKFTPNLS 403
Qy 132 VDVGFAHLRGKKHVFETQNIKG---LLLVADYTTKATANLYGLNLYRF 179
Db 404 VDLGAYLKGKKVHFKVEKVTIGDKRTLTNTTANYTSQAHANLYGLNLYSF 455

RESULT 12
```

```

Q9KHH3
ID Q9KHH3 PRELIMINARY: PRT: 455 AA.
AC Q9KHH3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC9795;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
vaccine candidate against experimental otitis media due to nontypeable
Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260352; AAF97566.1; -.
SQ SEQUENCE 455 AA: 49169 MW: 77A8BA97BCE2B450 CRC64;

Query Match 58.4%; Score 559; DB 2; Length 455;
Best Local Similarity 61.6%; Pred. No. 1.1e-40;
Matches 106; Conservative 29; Mismatches 31; Indels 6; Gaps 2;

QY 12 KGSLLTLKLPAYWELSGFHQLTQWATHYSKYTWSRFKELRGKYQDGSYEAFYKKEEY 71
DB 286 KGNLTLLPDYLELSGFHQLTKFAVHYSKYTHWSRLTKLHASFEDGK--KAFDKELQY 343
QY 72 KNSRFAICTTYSNLDAITLRLAGLADKASKTHLSASIPDTRMYSGATYKFTPNLS 131
DB 344 SNNRVALGASYNLDKLTLAGIADQAAASHRRSAIPDTRTWSLGATYKFTPNLS 403
QY 132 VDVGFAHLRGKKHFEVETQNIKG----LLIVEADYTTKATANLYGLNLNRYF 179
DB 404 VDLGYAYLKGKKVHFKEVKTIGDKRTLTLNTTANYTSQAHANLYGLNLNYSF 455

Query Match 58.4%; Score 559; DB 2; Length 455;
Best Local Similarity 61.6%; Pred. No. 1.1e-40;
Matches 106; Conservative 29; Mismatches 31; Indels 6; Gaps 2;

QY 12 KGSLLTLKLPAYWELSGFHQLTQWATHYSKYTWSRFKELRGKYQDGSYEAFYKKEEY 71
DB 286 KGNLTLLPDYLELSGFHQLTKFAVHYSKYTHWSRLTKLHASFEDGK--KAFDKELQY 343
QY 72 KNSRFAICTTYSNLDAITLRLAGLADKASKTHLSASIPDTRMYSGATYKFTPNLS 131
DB 344 SNNRVALGASYNLDKLTLAGIADQAAASHRRSAIPDTRTWSLGATYKFTPNLS 403
QY 132 VDVGFAHLRGKKHFEVETQNIKG----LLIVEADYTTKATANLYGLNLNRYF 179
DB 404 VDLGYAYLKGKKVHFKEVKTIGDKRTLTLNTTANYTSQAHANLYGLNLNYSF 455

RESULT 13
Q00450
ID Q00450 PRELIMINARY: PRT: 455 AA.
AC Q00450;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein p1 precursor, subtype 3L (OMP p1).
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1613 / SEROTYPE B;
RX MEDLINE=90035394; PubMed=2572549;
RA Munson R.S. Jr., Grass S., Einhorn M., Bailey C., Newell C.;
RT "Comparative analysis of the structures of the outer membrane protein
p1 genes from major clones of Haemophilus influenzae type b.";
RL Infect. Immun. 57:3300-3305(1989).
DR EMBL: AF260352; AAF97566.1; -.
SQ SEQUENCE FROM N.A.
RC STRAIN=7109, AND 6107;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
vaccine candidate against experimental otitis media due to nontypeable
Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).

```

```

CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: STRONG, TO E.COLI LONG-CHAIN FATTY ACID TRANSPORT
CC PROTEIN (FADL).
DR EMBL: M63151; AAK24991.1; -.
DR EMBL: AF260351; AAF97565.1; -.
DR EMBL: AF260350; AAF97564.1; -.
DR PIR: A40183; A40183.
KW Outer membrane; Signal; Transmembrane.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 455 OUTER MEMBRANE PROTEIN P1.
SQ SEQUENCE 455 AA: 49141 MW: 799059130A39D9E1 CRC64;

Query Match 58.4%; Score 559; DB 2; Length 455;
Best Local Similarity 61.6%; Pred. No. 1.1e-40;
Matches 106; Conservative 29; Mismatches 31; Indels 6; Gaps 2;

QY 12 KGSLLTLKLPAYWELSGFHQLTQWATHYSKYTWSRFKELRGKYQDGSYEAFYKKEEY 71
DB 286 KGNLTLLPDYLELSGFHQLTKFAVHYSKYTHWSRLTKLHASFEDGK--KAFDKELQY 343
QY 72 KNSRFAICTTYSNLDAITLRLAGLADKASKTHLSASIPDTRMYSGATYKFTPNLS 131
DB 344 SNNRVALGASYNLDKLTLAGIADQAAASHRRSAIPDTRTWSLGATYKFTPNLS 403
QY 132 VDVGFAHLRGKKHFEVETQNIKG----LLIVEADYTTKATANLYGLNLNRYF 179
DB 404 VDLGYAYLKGKKVHFKEVKTIGDKRTLTLNTTANYTSQAHANLYGLNLNYSF 455

RESULT 14
Q9KHH5
ID Q9KHH5 PRELIMINARY: PRT: 455 AA.
AC Q9KHH5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=7004;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
vaccine candidate against experimental otitis media due to nontypeable
Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260348; AAF97562.1; -.
SQ SEQUENCE 455 AA: 49165 MW: 280E211B16CA18C0 CRC64;

Query Match 58.2%; Score 557; DB 2; Length 455;
Best Local Similarity 61.6%; Pred. No. 1.1e-40;
Matches 106; Conservative 28; Mismatches 32; Indels 6; Gaps 2;

QY 12 KGSLLTLKLPAYWELSGFHQLTQWATHYSKYTWSRFKELRGKYQDGSYEAFYKKEEY 71
DB 286 KGNLTLLPDYLELSGFHQLTKFAVHYSKYTHWSRLTKLHASFEDGK--KAFDKELQY 343
QY 72 KNSRFAICTTYSNLDAITLRLAGLADKASKTHLSASIPDTRMYSGATYKFTPNLS 131
DB 344 SNNRVALGASYNLDKLTLAGIADQAAASHRRSAIPDTRTWSLGATYKFTPNLS 403
QY 132 VDVGFAHLRGKKHFEVETQNIKG----LLIVEADYTTKATANLYGLNLNRYF 179
DB 404 VDLGYAYLKGKKVHFKEVKTIGDKRTLTLNTTANYTSQAHANLYGLNLNYSF 455

RESULT 15

```

```

Q9KHF6      PRELIMINARY;      PRT;      451 AA.
ID Q9KHF6;
AC Q9KHF6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein P1.
GN OMPP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13-H-1157;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260373; AAF97589.1; -.
SQ SEQUENCE 451 AA; 48446 MW; F4DD7C9354DFCF13 CRC64;

Query Match      57.8%; Score 553.5; DB 2; Length 451;
Best Local Similarity 62.1%; Pred. No. 3.2e-40;
Matches 105; Conservative 27; Mismatches 34; Indels 3; Gaps 2;

QY 12 KGSITLKLPAWELSGFHQLTQWAIHYSKYKTEWSRFKELRGKYQDGSYGAEFTKKEEY 71
Db 285 KGDUTLTPDYLELSGFHQLTDKLVAVHSYKYTHWSRLTKLVASFEDGK--KAFDKELQY 342
QY 72 KDNSRFAIGTGYSLNDALTFRAGLAYDKAASKTHLSASIPDTRMWSYGATYKFTPNLS 131
Db 343 SNNRSVALGASYNLAELTLRAGIAYDQASRHQSRSAIPDTRTWISLGATYKFTPNLS 402
QY 132 VDVGFALHGRGKKHFVETQNIKGLLVEADYTTKATNLYGLNLNRYF 179
Db 403 VDLGVAYLKGKKVHFKEVQRAAGGHIITNTANYTSQAHANLYGLNLNYSF 451

RESULT 16
Q9KHF4      PRELIMINARY;      PRT;      450 AA.
ID Q9KHF4;
AC Q9KHF4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein P1.
GN OMPP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=7424;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260375; AAF97589.1; -.
SQ SEQUENCE 450 AA; 48593 MW; CF17201C42BDA757 CRC64;

Query Match      57.7%; Score 552.5; DB 2; Length 450;
Best Local Similarity 61.8%; Pred. No. 3.9e-40;
Matches 105; Conservative 28; Mismatches 34; Indels 3; Gaps 2;

QY 11 GK-GSLTLKLPAWELSGFHQLTQWAIHYSKYKTEWSRFKELRGKYQDGSYGAEFTKKE 69
Db 283 KGTGDLTFTLPDYLELSGFHQLTDKLVAVHSYKYTHWSRLTKLVASYENG--KAFDKEL 340
QY 70 EYKNSRFAIGTGYSLNDALTFRAGLAYDKAASKTHLSASIPDTRMWSYGATYKFTPN 129
Db 341 QYSNRSVALGASYNLYEKLTLRAGIAYDQASRHQSRSAIPDTRTWISLGATYKFTPN 400
QY 130 LSVGVGFALHGRGKKHFVETQNIKGLLVEADYTTKATNLYGLNLNRYF 179
Db 401 LSVDLGVAYLKGKKVHFKEVQRAAGGHIITNTANYTSQAHANLYGLNLNYSF 450

RESULT 17
Q9KHF3      PRELIMINARY;      PRT;      450 AA.
ID Q9KHF3;
AC Q9KHF3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein P1.
GN OMPP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4-H-1093;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260376; AAF97590.1; -.
SQ SEQUENCE 450 AA; 48741 MW; 3C68E144339F41CD CRC64;

Query Match      57.7%; Score 552.5; DB 2; Length 450;
Best Local Similarity 61.8%; Pred. No. 3.9e-40;
Matches 105; Conservative 28; Mismatches 34; Indels 3; Gaps 2;

QY 11 GK-GSLTLKLPAWELSGFHQLTQWAIHYSKYKTEWSRFKELRGKYQDGSYGAEFTKKE 69
Db 283 KGTGDLTFTLPDYLELSGFHQLTDKLVAVHSYKYTHWSRLTKLVASYENG--KAFDKEL 340
QY 70 EYKNSRFAIGTGYSLNDALTFRAGLAYDKAASKTHLSASIPDTRMWSYGATYKFTPN 129
Db 341 QYSNRSVALGASYNLYEKLTLRAGIAYDQASRHQSRSAIPDTRTWISLGATYKFTPN 400
QY 130 LSVGVGFALHGRGKKHFVETQNIKGLLVEADYTTKATNLYGLNLNRYF 179
Db 401 LSVDLGVAYLKGKKVHFKEVQRAAGGHIITNTANYTSQAHANLYGLNLNYSF 450

RESULT 18
Q9KHGO      PRELIMINARY;      PRT;      451 AA.
ID Q9KHGO;
AC Q9KHGO;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein P1.
GN OMPP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4-H-1094;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,

```



RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;  
RT "Variability of outer membrane protein P1 and its evaluation as a  
RT vaccine candidate against experimental otitis media due to nontypeable  
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";  
RL Infect. Immun. 68:4505-4517(2000).  
DR EMBL: AF260367; AAF97581.1; -.  
SQ SEQUENCE 451 AA; 48640 MW; 72C6977C6F2A7BA1 CRC64;

Query Match 57.5%; Score 550.5; DB 2; Length 451;  
Best Local Similarity 61.2%; Pred. No. 5.9e-40;  
Matches 109; Conservative 26; Mismatches 38; Indels 5; Gaps 3;  
QY 4 GVLGPIYIGKSLTLKPAYWELSGFHQLTQDQWAIHYSYKYTEWSRKFELRGKYQDGSYG 63  
DB 277 GVIGAG-KKGDLTLPDYLELGGPHQTLKLVAVHSYKYTHWSRUTKLNASPEDGK-K 333  
QY 64 AFTKKEEYKDNRSFAITGTTYSINDALTLAGLAYDKAAKTHLSASIPDTRMWSIGAT 123  
DB 334 AFDEKELQYSNRSVALGASVDLYEKLTLRAGIAYDQAASHRQSAIPDTRTWSLGAT 393  
QY 124 YKFTPNLSVDVGFALHGRKKHFEVETONI-KGLLLVEADYTTKATANLYGLNLNYSF 179  
DB 394 YKFTPNLSVDLGAYLKGKKVHFEVKTKTGVILNATADYTSQAHANLYGLNLNYSF 451

RESULT 19  
Q9KHG9 PRELIMINARY; PRT; 456 AA.  
AC Q9KHC9;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE Outer membrane protein P1.  
GN OMPPI.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1071;  
RX MEDLINE=20359342; PubMed=10899849;  
RA Bouduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,  
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;  
RT "Variability of outer membrane protein P1 and its evaluation as a  
RT vaccine candidate against experimental otitis media due to nontypeable  
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";  
RL Infect. Immun. 68:4505-4517(2000).  
DR EMBL: AF260358; AAF97572.1; -.  
SQ SEQUENCE 456 AA; 49130 MW; C16D7F3809AC8FA2 CRC64;

Query Match 57.4%; Score 549.5; DB 2; Length 456;  
Best Local Similarity 61.3%; Pred. No. 7.3e-40;  
Matches 111; Conservative 28; Mismatches 33; Indels 9; Gaps 5;  
QY 4 GVLGPIYIGK-GSITLKPAYWELSGFHQLTQDQWAIHYSYKYTEWSRKFELRGKYQDGSYG 62  
DB 280 GVIG--AGTKGNLTLPDYLELGGPHQTLKLVAVHSYKYTHWSRUTKLNASPEDGK-- 335  
QY 63 EAFKKEEYKDNRSFAITGTTYSINDALTLAGLAYDKAAKTHLSASIPDTRMWSIGAT 122  
DB 336 KAFDEKELQYSNRSVALGASYNLYEKLTLRAGIAYDQAASHRQSAIPDTRTWSLGAT 395  
QY 123 YKFTPNLSVDVGFALHGRKKHFEVETONI-KGLLLVE-ADYTTKATANLYGLNLNYSF 178  
DB 396 YKFTPNLSVDLGAYLKGKKVHFEVKTKTGVILNATADYTSQAHANLYGLNLNYS 455

Query Match 57.4%; Score 549.5; DB 2; Length 456;  
Best Local Similarity 61.3%; Pred. No. 7.3e-40;  
Matches 111; Conservative 28; Mismatches 33; Indels 9; Gaps 5;  
QY 4 GVLGPIYIGK-GSITLKPAYWELSGFHQLTQDQWAIHYSYKYTEWSRKFELRGKYQDGSYG 62  
DB 280 GVIG--AGTKGNLTLPDYLELGGPHQTLKLVAVHSYKYTHWSRUTKLNASPEDGK-- 335  
QY 63 EAFKKEEYKDNRSFAITGTTYSINDALTLAGLAYDKAAKTHLSASIPDTRMWSIGAT 122  
DB 336 KAFDEKELQYSNRSVALGASYNLYEKLTLRAGIAYDQAASHRQSAIPDTRTWSLGAT 395  
QY 123 YKFTPNLSVDVGFALHGRKKHFEVETONI-KGLLLVE-ADYTTKATANLYGLNLNYSF 178  
DB 396 YKFTPNLSVDLGAYLKGKKVHFEVKTKTGVILNATADYTSQAHANLYGLNLNYS 455

QY 179 F 179  
DB 456 F 456

RESULT 20

Q9KHH6 PRELIMINARY; PRT; 456 AA.  
AC Q9KHH6;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE Outer membrane protein P1.  
GN OMPPI.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=7416;  
RX MEDLINE=20359342; PubMed=10899849;  
RA Bouduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,  
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;  
RT "Variability of outer membrane protein P1 and its evaluation as a  
RT vaccine candidate against experimental otitis media due to nontypeable  
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";  
RL Infect. Immun. 68:4505-4517(2000).  
DR EMBL: AF260347; AAF97561.1; -.  
SQ SEQUENCE 456 AA; 49370 MW; 0FA50D60942905D0 CRC64;

Query Match 57.4%; Score 549; DB 2; Length 456;  
Best Local Similarity 61.6%; Pred. No. 8.1e-40;  
Matches 106; Conservative 28; Mismatches 32; Indels 6; Gaps 3;  
QY 12 KGSITLKPAYWELSGFHQLTQDQWAIHYSYKYTEWSRKFELRGKYQDGSYGAEFTKKEEY 71  
DB 287 KGDLTLTLPDYLELGGPHQTLKLVAVHSYKYTHWSRUTKLNASPEDGK--KAPDKELQY 344  
QY 72 KDNRSFAITGTTYSINDALTLAGLAYDKAAKTHLSASIPDTRMWSIGATYKFTPNLS 131  
DB 345 SNRSVALGASYNLYEKLTLRAGIAYDQAASHRQSAIPDTRTWSLGATYKFTPNLS 404  
QY 132 VDVGFALHGRKKHFEVETONI-KGLLL-VEADYTTKATANLYGLNLNYSF 179  
DB 405 VDLGAYLKGKKVHFEVKTKTGVILNATADYTSQAHANLYGLNLNYSF 456

RESULT 21  
Q9KHG7 PRELIMINARY; PRT; 456 AA.  
AC Q9KHG7;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE Outer membrane protein P1.  
GN OMPPI.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LO-127;  
RX MEDLINE=20359342; PubMed=10899849;  
RA Bouduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,  
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;  
RT "Variability of outer membrane protein P1 and its evaluation as a  
RT vaccine candidate against experimental otitis media due to nontypeable  
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";  
RL Infect. Immun. 68:4505-4517(2000).  
DR EMBL: AF260360; AAF97574.1; -.  
SQ SEQUENCE 456 AA; 49220 MW; 7AC1F56499D94C91 CRC64;

Query Match 57.4%; Score 549; DB 2; Length 456;  
Best Local Similarity 62.2%; Pred. No. 8.1e-40;  
Matches 107; Conservative 25; Mismatches 34; Indels 6; Gaps 3;

QY 12 KGSITLKPAYWELSGFHQLTQDQWAIHYSYKYTEWSRKFELRGKYQDGSYGAEFTKKEEY 71

```

Db 287 KGDLTLTPDYLELGSFQHLTKLAVHSYKYTHWSRLTKLHSPEDGK--KAFDKELQY 344
Qy 72 KNSRFAIGTYSYSLNDALTLRAGLAYDKAASKTHLSASIPDTRMWSYIGATYKFTPNLS 131
Db 345 SNNSRVALGASYNLYEKLTLRAGIAYDQAASRHSRAAIPDTRTWYSLGATYKFTPNLS 404
Qy 132 VDVGFALHGRGKKHVFETONI---KGL-LLVEADYTTKATANLYGLNLYRF 179
Db 405 VDLGVAYLKGKKVHFKEVNTIGKESGLPLTTTANTYTSOAHANLYGLNLYSF 456

RESULT 22
Q9KHG2 PRELIMINARY; PRT; 451 AA.
AC Q9KHG2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=6181;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL; AF260365; AAF97579.1; -
SQ SEQUENCE 451 AA; 48611 MW; B616AAB6FEF0252A CRC64;

Query Match 57.2%; Score 547; DB 2; Length 451;
Best Local Similarity 62.4%; Pred. No. 1.2e-39;
Matches 106; Conservative 27; Mismatches 33; Indels 4; Gaps 3;

Qy 12 KGSLLTKLPAYWELSGFQHLTDQMAIHYSKYKTEWSRFEKLRGKYQDGSYGAEFTKKEEY 71
Db 284 KGDLTLTPDYLELGSFQHLTKLAVHSYKYTHWSRLTKLYASSENK--KAFDKELQY 341
Qy 72 KNSRFAIGTYSYSLNDALTLRAGLAYDKAASKTHLSASIPDTRMWSYIGATYKFTPNLS 131
Db 342 SNNSRVALGASYNLYEKLTLRAGIAYDQAASRHSRAAIPDTRTWYSLGATYKFTPNLS 401
Qy 132 VDVGFALHGRGKKHVFETONI-KGLLL-VEADYTTKATANLYGLNLYRF 179
Db 402 VDLGVAYLKGKKVHFKEVATIGKSVILNTTANTYTSOAHANLYGLNLYSF 451

```

```

RESULT 23
Q9K324 PRELIMINARY; PRT; 459 AA.
AC Q9K324;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=218, AND 200;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,

```

```

RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL; AF260344; AAF97558.1; -
SQ SEQUENCE 459 AA; 49717 MW; 54490C08E4729870 CRC64;

Query Match 57.2%; Score 547; DB 2; Length 459;
Best Local Similarity 61.6%; Pred. No. 1.2e-39;
Matches 106; Conservative 27; Mismatches 33; Indels 6; Gaps 3;

Qy 12 KGSLLTKLPAYWELSGFQHLTDQMAIHYSKYKTEWSRFEKLRGKYQDGSYGAEFTKKEEY 71
Db 290 KGDLTLTPDYLELGSFQHLTKLAVHSYKYTHWSRLTKLNASFEDGK--KAFDKELQY 347
Qy 72 KNSRFAIGTYSYSLNDALTLRAGLAYDKAASKTHLSASIPDTRMWSYIGATYKFTPNLS 131
Db 348 SNNSRVALGASYNLYEKLTLRAGIAYDQAASRHSRAAIPDTRTWYSLGATYKFTPNLS 407
Qy 132 VDVGFALHGRGKKHVFETONI---KGLLL-VEADYTTKATANLYGLNLYRF 179
Db 408 VDLGVAYLKGKKVHFKEVKTIGDERSLALNTTANTYTSOAHANLYGLNLYSF 459

RESULT 24
Q9KHH8 PRELIMINARY; PRT; 459 AA.
AC Q9KHH8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=667;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL; AF260345; AAF97559.1; -
SQ SEQUENCE 459 AA; 49773 MW; 5E6DCD78F9A8FF8A CRC64;

Query Match 57.2%; Score 547; DB 2; Length 459;
Best Local Similarity 61.6%; Pred. No. 1.2e-39;
Matches 106; Conservative 27; Mismatches 33; Indels 6; Gaps 3;

Qy 12 KGSLLTKLPAYWELSGFQHLTDQMAIHYSKYKTEWSRFEKLRGKYQDGSYGAEFTKKEEY 71
Db 290 KGDLTLTPDYLELGSFQHLTKLAVHSYKYTHWSRLTKLNASFEDGK--KAFDKELQY 347
Qy 72 KNSRFAIGTYSYSLNDALTLRAGLAYDKAASKTHLSASIPDTRMWSYIGATYKFTPNLS 131
Db 348 SNNSRVALGASYNLYEKLTLRAGIAYDQAASRHSRAAIPDTRTWYSLGATYKFTPNLS 407
Qy 132 VDVGFALHGRGKKHVFETONI---KGLLL-VEADYTTKATANLYGLNLYRF 179
Db 408 VDLGVAYLKGKKVHFKEVKTIGDERSLALNTTANTYTSOAHANLYGLNLYSF 459

RESULT 25
Q9KHH7 PRELIMINARY; PRT; 459 AA.
AC Q9KHH7;

```

```
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-DEC-2001 (TREMBlrel. 15, Last sequence update)
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=199;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL; AF260346; AAF97560.1; -.
SQ SEQUENCE 459 AA; 49745 MW; 5A46FF07187E98B3 CRC64;

Query Match 57.2%; Score 547; DB 2; Length 459;
Best Local Similarity 61.6%; Pred. No. 1.2e-39;
Matches 106; Conservative 27; Mismatches 33; Indels 6; Gaps 3;

QY 12 KGSLLTKLPAYWELSGFHLTDQWAIHYSYKYTEWSRFKELRGKYQDGSYEAFKKKEEY 71
DB 290 KGDLTTLTPDYLELSGFHLTDKLAIVHYSKYTHWSRLTKLNASFEDGK--KAFDKELQY 347

QY 72 KNSRFAIGTTYSLNDALTLRAGLAYDKAASKTHLSASIPDTRMWSIGATYKFTPNLS 131
DB 348 SNNRVALGASYNLYEKLTLRAGIAYDQAAASHRQSAIIPDTRTWYSLGATYKFTPNLS 407

QY 132 VDVGFAHLRGKKKHVFETONI---KGLLL-VEADYTTKATANLYGLNLNRYF 179
DB 408 VDLGYALYLGKKVHKVFEVKTIGDRSLALNTTANTYTSQAHANLYGLNLNYSF 459

RESULT 26
Q9KHG8 PRELIMINARY; PRT; 456 AA.
AC Q9KHG8;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BCH-2;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL; AF260359; AAF97573.1; -.
SQ SEQUENCE 456 AA; 49075 MW; CF7B3ECE4213C841 CRC64;

Query Match 57.0%; Score 545.5; DB 2; Length 456;
Best Local Similarity 60.2%; Pred. No. 1.6e-39;
Matches 109; Conservative 29; Mismatches 34; Indels 9; Gaps 5;

QY 4 GVLGPIYIK-GSLTLKLPAYWELSGFHLTDQWAIHYSYKYTEWSRFKELRGKYQDGSY 62
DB 280 GVIG--AGTGNLTFTLPDYLELSGFHLTDKLAIVHYSKYTHWSRLTKLNASFEDGK-- 335
```

```
QY 63 EAFTKKEEYKDNSRFAIGTTYSLNDALTLRAGLAYDKAASKTHLSASIPDTRMWSIG 122
DB 336 KAFDKELQYSSNRSVALGASYNLYEKLTLRAGIAYDQAAASHRQSAIIPDTRTWYSLG 395

QY 123 TYKFTPNLSVDVGFAHLRGKKKHVFETONI---KGL-LLVEADYTTKATANLYGLNLN 178
DB 396 TYKFTPNLSVDVGAYLKGKVKHFEVKTIGKSGLPVTATANTYTSQAHANLYGLNLNYS 455

QY 179 F 179
DB 456 F 456

RESULT 27
Q9KH14 PRELIMINARY; PRT; 456 AA.
AC Q9KH14;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BCH-1;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL; AF260337; AAF97551.1; -.
SQ SEQUENCE 456 AA; 49402 MW; 8CF3E87CB81DD6E6 CRC64;

Query Match 56.8%; Score 544; DB 2; Length 456;
Best Local Similarity 61.4%; Pred. No. 2.2e-39;
Matches 105; Conservative 28; Mismatches 32; Indels 6; Gaps 3;

QY 13 GSLLTKLPAYWELSGFHLTDQWAIHYSYKYTEWSRFKELRGKYQDGSYEAFKKKEEY 72
DB 288 GDLTTLTPDYLELSGFHLTDKLAIVHYSKYTHWSRLTKLNASFEDGK--KAFDKELQYS 345

QY 73 DNSRFAIGTTYSLNDALTLRAGLAYDKAASKTHLSASIPDTRMWSIGATYKFTPNLSV 132
DB 346 NNSRVALGASYNLYEKLTLRAGIAYDQAAASHRQSAIIPDTRTWYSLGATYKFTPNLSV 405

QY 133 DVGFAHLRGKKKHVFETONI---KGLLL-VEADYTTKATANLYGLNLNRYF 179
DB 406 DLGYALYLGKKVHKVFEVKTIGDRSLALNTTANTYTSQAHANLYGLNLNYSF 456

RESULT 28
Q9KH12 PRELIMINARY; PRT; 456 AA.
AC Q9KH12;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3-H-1552;
RX MEDLINE=20359342; PubMed=10899849;
```

RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,  
RA "Variability of outer membrane protein P1 and its evaluation as a  
RT vaccine candidate against experimental otitis media due to nontypeable  
RT Haemophilus influenzae: an unambiguous, multifaceted approach."  
RL Infect. Immun. 68:4505-4517(2000).  
DR EMBL: AF260339; AAF97553.1; -.  
SQ SEQUENCE 456 AA; 49403 MW; F95A977CBCE4F057 CRC64;

Query Match 56.8%; Score 544; DB 2; Length 456;  
Best Local Similarity 61.4%; Pred. No. 2.2e-39;  
Matches 105; Conservative 28; Mismatches 32; Indels 6; Gaps 3;

Qy 13 GSLTLKLPAYWELSGFHQTLTDQWAIHYSKYTWSRFKELRGKYQDGSYEAFTKKEEY 72

Db 288 GDLTLTPDYLELSGFHQLTDKLVHYSKYTHWSRLTKLNASFEDGK--KAFDKELOYS 345

Qy 73 DNSRFAIGTYYSLNDALTLAGLAYDKAASKTHLSASIPDTRMWSYIGATYKFTPNLSV 132

Db 346 NNSRVALGASYNLEKLTLAGIAYDQAAHRHQRSAIPDTRTWYSLGATYKFTPNLSV 405

Qy 133 DVGFALRGKKKHFEVETQNI---KGLLL-VEADYTTKATANLYGLNLYRF 179

Db 406 DLGYALGKKVHKVEKVTIGDKRSLALNTTANTYTSQAHANLYGLNLYSF 456

RESULT 29

Q9KHG6 PRELIMINARY; PRT; 457 AA.

AC Q9KHG6;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE Outer membrane protein P1.

GN OMPP1.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Haemophilus.

OX NCBI\_TaxID=727;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=5;

RX MEDLINE=20359342; PubMed=10899849;

RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,

RA "Variability of outer membrane protein P1 and its evaluation as a

RT vaccine candidate against experimental otitis media due to nontypeable

RT Haemophilus influenzae: an unambiguous, multifaceted approach."

RL Infect. Immun. 68:4505-4517(2000).

DR EMBL: AF260361; AAF97575.1; -.  
SQ SEQUENCE 457 AA; 49292 MW; D25CB4D6E81A965F CRC64;

Query Match 56.7%; Score 542.5; DB 2; Length 457;  
Best Local Similarity 60.6%; Pred. No. 3e-39;  
Matches 109; Conservative 28; Mismatches 36; Indels 7; Gaps 4;

Qy 4 GVLPGYIGKSLTKLPAYWELSGFHQTLTDQWAIHYSKYTWSRFKELRGKYQDGSYE 63

Db 281 GVIGAG-KKGDLTLTPDYLELSGFHQLTDKLVHYSKYTHWSRLTKLNANFEDGK--K 337

Qy 64 AFTKKEEKDNSRFAIGTYYSLNDALTLAGLAYDKAASKTHLSASIPDTRMWSYIGAT 123

Db 338 AFDKELOYSNNSRVALGASYNLEKLTLAGIAYDQAAHRHQRSAIPDTRTWYSLGAT 397

Qy 124 YKFTPNLSVDVGFALRGKKKHFEVETQNI---KGLLLV-ADYTTKATANLYGLNLYRF 179

Db 398 YKFTPNLSVDLGYALGKKVHKVEKVTIGDKRSLALNTTANTYTSQAHANLYGLNLYSF 457

RESULT 30

Q9KHH9 PRELIMINARY; PRT; 456 AA.

ID Q9KHH9

AC Q9KHH9;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE Outer membrane protein P1.  
GN OMPP1.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=27-H-1433;  
RX MEDLINE=20359342; PubMed=10899849;  
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,  
RA "Variability of outer membrane protein P1 and its evaluation as a  
RT vaccine candidate against experimental otitis media due to nontypeable  
RT Haemophilus influenzae: an unambiguous, multifaceted approach."  
RL Infect. Immun. 68:4505-4517(2000).  
DR EMBL: AF260342; AAF97556.1; -.  
SQ SEQUENCE 456 AA; 49239 MW; 35101EFC7A6769D2 CRC64;

Query Match 56.2%; Score 538; DB 2; Length 456;  
Best Local Similarity 61.0%; Pred. No. 7.4e-39;  
Matches 105; Conservative 27; Mismatches 34; Indels 6; Gaps 3;

Qy 12 KGSUTLKLPAWELSGFHQTLTDQWAIHYSKYTWSRFKELRGKYQDGSYEAFTKKEEY 71

Db 287 KGDLTLTPDYLELSGFHQLTDKLVHYSKYTHWSRLTRLYASSENGK--KAFDKELOY 344

Qy 72 KDNSRFAIGTYYSLNDALTLAGLAYDKAASKTHLSASIPDTRMWSYIGATYKFTPNLS 131

Db 345 SNNRVALGASYNLEKLTLAGIAYDQAAHRHQRSAIPDTRTWYSLGATYKFTPNLS 404

Qy 132 VDVGFAHLRGKKKHFEVETQNI---KGLLL-VEADYTTKATANLYGLNLYRF 179

Db 405 VDLGYALGKKVHKVEKVTIGDKRSLALNTTANTYTSQAHANLYGLNLYSF 456

RESULT 31

Q9K397

ID Q9K397 PRELIMINARY; PRT; 458 AA.

AC Q9K397;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE Outer membrane protein P1.

GN OMPP1.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Haemophilus.

OX NCBI\_TaxID=727;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20359342; PubMed=10899849;

RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,

RA "Variability of outer membrane protein P1 and its evaluation as a

RT vaccine candidate against experimental otitis media due to nontypeable

RT Haemophilus influenzae: an unambiguous, multifaceted approach."

RL Infect. Immun. 68:4505-4517(2000).

DR EMBL: AF260356; AAF97570.1; -.  
DR EMBL: AF260355; AAF97569.1; -.  
SQ SEQUENCE 458 AA; 49674 MW; 1F886C1FAD682F94 CRC64;

Query Match 56.2%; Score 538; DB 2; Length 458;  
Best Local Similarity 60.5%; Pred. No. 7.4e-39;  
Matches 104; Conservative 27; Mismatches 35; Indels 6; Gaps 3;

Qy 12 KGSUTLKLPAWELSGFHQTLTDQWAIHYSKYTWSRFKELRGKYQDGSYEAFTKKEEY 71

Db 289 KGDLTLTPDYLELSGFHQLTDKLVHYSKYTHWSRLTRKLUHASFEDGK--KAFDKELOY 346



```
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22-H-1154;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RA "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260366; AAF97580.1; -.
SQ SEQUENCE 454 AA; 49986 MW; 494FBAC2600DA2C8 CRC64;

Query Match 55.7%; Score 533; DB 2; Length 454;
Best Local Similarity 61.0%; Pred. No. 2e-38;
Matches 105; Conservative 26; Mismatches 35; Indels 6; Gaps 3;

QY 12 KGSITLKLPAWELSGFHQLTDQWAIHYSKYKTYTWSRFKELRGYQDGSYGAEATKKEEY 71
Db 285 KGDTLTPDYLSLGSFHQLTKLAVHSYKYTHWSRLTRLYASSENGK--KAFDKELQY 342
QY 72 KDNSRFAIGTYSLNDALTLAGLAYDKAASKTHLSASIPDTRMWSYSGATYKFTPNLS 131
Db 343 SNNRVALGASYNLYEKLTLRAGIAYDQASRHORSRAIPTDRTWISLGTATYKFTPNLS 402
QY 132 VDVGFAHLRGKKKHVFETQNI---KGLLL-VEADYTTKATANLYGLNLNRYF 179
Db 403 VDLGYAYLKGKKVHFKEVKTIGDARSALNTTANYTSQAHANLYGLNLNYSF 454

RESULT 36
QY 12 KGSITLKLPAWELSGFHQLTDQWAIHYSKYKTYTWSRFKELRGYQDGSYGAEATKKEEY 71
Db 285 KGDTLTPDYLSLGSFHQLTKLAVHSYKYTHWSRLTRLYASSENGK--KAFDKELQY 342
QY 72 KDNSRFAIGTYSLNDALTLAGLAYDKAASKTHLSASIPDTRMWSYSGATYKFTPNLS 131
Db 343 SNNRVALGASYNLYEKLTLRAGIAYDQASRHORSRAIPTDRTWISLGTATYKFTPNLS 402
QY 132 VDVGFAHLRGKKKHVFETQNI---KGLLL-VEADYTTKATANLYGLNLNRYF 179
Db 403 VDLGYAYLKGKKVHFKEVKTIGDARSALNTTANYTSQAHANLYGLNLNYSF 454

RESULT 37
QY 132 VDVGFAHLRGKKKHVFETQNI---KGLLL-VEADYTTKATANLYGLNLNRYF 179
Db 405 VDLGYAYLKGKKVHFKEVKTIGDARSALNTTANYTSQAHANLYGLNLNYSF 456

RESULT 38
QY 12 KGSITLKLPAWELSGFHQLTDQWAIHYSKYKTYTWSRFKELRGYQDGSYGAEATKKEEY 71
Db 291 KADLTFTLPDYLSLGSFHQLTKLAVHSYKYTHWSRLTKLHASFEDGK--KAFDKELQY 348
QY 72 KDNSRFAIGTYSLNDALTLAGLAYDKAASKTHLSASIPDTRMWSYSGATYKFTPNLS 131
Db 349 SNNRVALGASYNLYEKLTLRAGIAYDQASRHORSRAIPTDRTWISLGTATYKFTPNLS 408
QY 132 VDVGFAHLRGKKKHVFETQNI---KGLLL-VEADYTTKATANLYGLNLNRYF 179
Db 409 VDLGYAYLKGKKVHFKEVKTIGEQRSLSLTFTTANYTSQAHANLYGLNLNYSF 460

Query Match 55.6%; Score 532; DB 2; Length 460;
Best Local Similarity 59.9%; Pred. No. 2.5e-38;
Matches 103; Conservative 27; Mismatches 36; Indels 6; Gaps 3;
```

```
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9-H-1194;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RA "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260357; AAF97571.1; -.
SQ SEQUENCE 460 AA; 49719 MW; F1CEBEB597A5CBA7 CRC64;

Query Match 55.6%; Score 532; DB 2; Length 460;
Best Local Similarity 59.9%; Pred. No. 2.5e-38;
Matches 103; Conservative 27; Mismatches 36; Indels 6; Gaps 3;

QY 12 KGSITLKLPAWELSGFHQLTDQWAIHYSKYKTYTWSRFKELRGYQDGSYGAEATKKEEY 71
Db 291 KADLTFTLPDYLSLGSFHQLTKLAVHSYKYTHWSRLTKLHASFEDGK--KAFDKELQY 348
QY 72 KDNSRFAIGTYSLNDALTLAGLAYDKAASKTHLSASIPDTRMWSYSGATYKFTPNLS 131
Db 349 SNNRVALGASYNLYEKLTLRAGIAYDQASRHORSRAIPTDRTWISLGTATYKFTPNLS 408
QY 132 VDVGFAHLRGKKKHVFETQNI---KGLLL-VEADYTTKATANLYGLNLNRYF 179
Db 409 VDLGYAYLKGKKVHFKEVKTIGEQRSLSLTFTTANYTSQAHANLYGLNLNYSF 460

RESULT 38
QY 12 KGSITLKLPAWELSGFHQLTDQWAIHYSKYKTYTWSRFKELRGYQDGSYGAEATKKEEY 71
Db 291 KADLTFTLPDYLSLGSFHQLTKLAVHSYKYTHWSRLTKLHASFEDGK--KAFDKELQY 348
QY 72 KDNSRFAIGTYSLNDALTLAGLAYDKAASKTHLSASIPDTRMWSYSGATYKFTPNLS 131
Db 349 SNNRVALGASYNLYEKLTLRAGIAYDQASRHORSRAIPTDRTWISLGTATYKFTPNLS 408
QY 132 VDVGFAHLRGKKKHVFETQNI---KGLLL-VEADYTTKATANLYGLNLNRYF 179
Db 409 VDLGYAYLKGKKVHFKEVKTIGEQRSLSLTFTTANYTSQAHANLYGLNLNYSF 460

Query Match 55.7%; Score 533; DB 2; Length 456;
Best Local Similarity 61.0%; Pred. No. 2e-38;
Matches 105; Conservative 26; Mismatches 35; Indels 6; Gaps 3;
```

```

SQ  SEQUENCE 456 AA; 49237 MW; 94A04FOA9A649750 CRC64;
Query Match 54.0%; Score 517; DB 2; Length 456;
Best Local Similarity 60.5%; Pred. No. 5e-37;
Matches 104; Conservative 24; Mismatches 38; Indels 6; Gaps 3;

OY 12 KGSLLTLKPAYWELSGFHQLTQDWAHYSKYKTEWSRFKELRGKYQDGSYEAFTKKEEY 71
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 287 KGLDTLTPDYLELSGFHQLTKLAVHYSKYKTHWSRLTRLYASSENGK--KAPDKELQY 344

OY 72 KNSRFAIGTYSLNDALTLRAGLAYDKAASKTHLSASIPDTRDMWYSIGATYKFFPNLS 131
   :|||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 345 SNNRSVALGASNYLYEKLTLRAGIAKDOAAASCLQRSAAIPDTRTWYSLGATYKFFPNLS 404

OY 132 VDVGFAHLRCKKHFVETQNI---KGLLL-VEADYTTKATANLYGLNLNRYF 179
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 405 VDLGYAYLKGKKVHFVEVKTIGDARSALNTTANYTFOAHANLYGLNLNRYF 456

RESULT 39
O9KL84 PRELIMINARY; PRT; 412 AA.
AC O9KL84;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Long-chain fatty acid transport protein.
GN VCA0862.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_Taxid=666;
RN [1]

SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
DR EMBL; AE004414; AAF96760.1;
DR TIGR; VCA0862;
DR InterPro; IPR001064; Crystallin.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 412 AA; 44784 MW; FB606C929838D891 CRC64;

Query Match 44.0%; Score 421; DB 16; Length 412;
Best Local Similarity 46.9%; Pred. No. 1.1e-28;
Matches 83; Conservative 27; Mismatches 61; Indels 6; Gaps 2;

OY 3 NGVLGPYIGKSLTLKPAYWELSGFHQLTQDWAHYSKYKTEWSRFKELRGKYQDGSY 62
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 SGIATSAVVPQKLIELPAIWELSGFHQLEQNAVHYSYQOTDWSFEELTATSSQCKNG 300

OY 63 EATKKEEYKDNRSFAIGTYSLNDALTLRAGLAYDKAASKTHLSASIPDTRDMWYSIGA 122
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 TCFQKIEQYQDNGRWSVGATYTLTTTLRAGLAFADEQAGKATL--SIPDSDFWISAGL 358

OY 123 TYKFTPNLSVDVGFALHGRKKHFEVETQNIKGLLLVEADYTTKATANLYGLNLNRYF 179
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 359 TYTLTLPOLTHADGALVQSRKGSFTKQNIQGNLQFESE----AVAYISALQFNRYF 411

RESULT 40
O82D48 PRELIMINARY; PRT; 423 AA.
ID O82D48
AC O82D48;
```

```

DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative long-chain fatty acid transport protein.
GN FADL_OR_YPO2744.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_Taxid=632;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feilwell T., Hamlin N., Holroyd S., Jagels K., Karlysheva I.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
DR EMBL; AJ414153; CAC92983.1;
DR InterPro; IPR001064; Crystallin.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 423 AA; 45668 MW; 8D11F6495FDB9FE9 CRC64;

Query Match 38.9%; Score 372.5; DB 16; Length 423;
Best Local Similarity 44.8%; Pred. No. 1.9e-24;
Matches 79; Conservative 23; Mismatches 70; Indels 5; Gaps 3;

OY 4 GVLGPYIGKSLTLKPAYWELSGFHQLTQDWAHYSKYKTEWSRFKELRGKYQDGSY 63
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 251 GGLGKRTVPGLSLTLNLPVWVSGYKNKVPQWAIHYSMAVTTWSSFKELKATASNGD--V 308

OY 64 AFTKKEEYKDNRSFAIGTYSLNDALTLRAGLAYDKAASKT-HLSASIPDTRDMWYSIGA 122
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 309 LFDKHEGFDAYRIALCTTYVDDNMWTFRTGIAFDSPIPAGNRSISIPDQDFWLSAGT 368

OY 123 TYKFTPNLSVDVGFALHGRKKHFEVETQNIKGLLLVEADYTTKATANLYGLNLNRYF 179
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 369 TYAFNKNASVDVGIAVMKGNVSITEKTPAPSNITYE--FNSKGSAMLYGVNFNTF 423

Search completed: May 12, 2003, 09:41:14
Job time : 32 secs
```





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 09:41:45 ; Search time 36 Seconds  
(without alignments)  
662.552 Million cell updates/sec

Title: US-09-995-493-52

Perfect score: 179

Sequence: 1 QHNGVLGPYIGKSLTKLP.....ADYTTKATANLYGLNLNRYR 179

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A\_Geneseq\_101002.\*

1: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13	7.3	77	AA12456	H.influenzae stral
2	10	5.6	35	AA140060	Hib OMP pl peptide
3	10	5.6	36	AA12455	HIBP1-14(400-437)
4	10	5.6	36	AA140066	Hib OMP pl peptide
5	10	5.6	39	AA12454	HIBP1-13(400-437)
6	10	5.6	39	AA140065	Hib OMP pl peptide
7	10	5.6	53	AA140100	Hib OMP pl peptide
8	10	5.6	54	AA140101	Hib OMP pl-P2 hybr
9	10	5.6	55	AA140099	Hib OMP pl-P2 hybr
10	8	4.5	636	AA196665	Putative P. abyss

11	7	3.9	61	20	AA133535	Tobacco DNA bindin
12	7	3.9	68	20	AA131478	A. thaliana RAP2.4
13	7	3.9	68	21	AA125837	AP2 domain amino a
14	7	3.9	69	20	AA131468	N. tabacum EREBP-2
15	7	3.9	69	21	AA125827	AP2 domain amino a
16	7	3.9	79	21	AA132799	zeamays protein f
17	7	3.9	93	22	AA132798	C glutamicum prote
18	7	3.9	95	21	AA132798	zeamays protein f
19	7	3.9	109	22	AA154058	propionibacterium
20	7	3.9	113	22	AA148379	propionibacterium
21	7	3.9	127	23	AA125758	Streptococcus poly
22	7	3.9	134	21	AA147231	Arabidopsis thalia
23	7	3.9	206	19	AA160957	Streptococcus pneu
24	7	3.9	207	23	AA126302	Streptococcus poly
25	7	3.9	218	22	AA102538	A. thaliana transc
26	7	3.9	219	21	AA132762	Eucalyptus grandis
27	7	3.9	230	20	AA121977	Sark protein fragm
28	7	3.9	313	21	AA138454	Arabidopsis thalia
29	7	3.9	313	21	AA138787	Arabidopsis thalia
30	7	3.9	330	21	AA138453	Arabidopsis thalia
31	7	3.9	330	21	AA138786	Arabidopsis thalia
32	7	3.9	334	21	AA138452	Arabidopsis thalia
33	7	3.9	334	21	AA138785	Arabidopsis thalia
34	7	3.9	338	22	AA171496	Human oilfactory re
35	7	3.9	338	22	AA172390	Human OR-like poly
36	7	3.9	466	18	AA117820	Pseudomonas putida
37	7	3.9	511	18	AA155454	H. pylori ORF 02ge
38	7	3.9	511	19	AA198255	H. pylori GHPO 155
39	7	3.9	511	19	AA171519	Helicobacter polyp
40	7	3.9	511	20	AA17176	H. pylori outer me
41	7	3.9	511	22	AA16315	H. pylori HPS143 p
42	7	3.9	511	22	AA16364	H. pylori HPC143 p
43	7	3.9	523	18	AA155286	H. pylori ORF 13ee
44	7	3.9	795	22	AA190056	C. glutamicum prote
45	7	3.9	865	22	AA198884	E. coli growth and
46	7	3.9	904	20	AA120767	Senescence-associ
47	6	3.4	8	19	AA181976	Bioactive fragment
48	6	3.4	15	14	AA142705	Human TGase3 50 kD
49	6	3.4	15	14	AA142705	Murine TGase3 50 k
50	6	3.4	15	16	AA182922	Japanese cedar pol
51	6	3.4	15	19	AA157752	Residues 146-160 o
52	6	3.4	15	19	AA157752	Residues 151-165 o
53	6	3.4	18	20	AA131457	A. thaliana AP2 do
54	6	3.4	18	21	AA125816	AP2 repeat AP2-R2
55	6	3.4	20	15	AA145556	Cry j I pollen all
56	6	3.4	20	15	AA145557	Cry j I pollen all
57	6	3.4	20	16	AA182505	Cry j I Japanese C
58	6	3.4	20	16	AA182506	Cry j I Japanese C
59	6	3.4	20	19	AA182575	Seq ID 10 from WO
60	6	3.4	22	18	AA108416	Exosite III from h
61	6	3.4	24	20	AA136717	Fragment of human
62	6	3.4	26	18	AA126439	Erythropoietin rec
63	6	3.4	30	15	AA145583	Cry j I pollen all
64	6	3.4	31	14	AA139868	C peptide RV-cl, r
65	6	3.4	33	12	AA12452	HIBP1-10(339-370)
66	6	3.4	33	14	AA140062	Hib OMP pl peptide
67	6	3.4	34	22	AA130954	Peptide #3605 enco
68	6	3.4	34	22	AA136145	Peptide #3651 enco
69	6	3.4	34	22	AA121531	Protein #3530 enco
70	6	3.4	34	22	AA156929	Human bone marrow
71	6	3.4	34	22	AA169315	Human bone marrow
72	6	3.4	34	22	AA117143	Peptide #3577 enco
73	6	3.4	34	22	AA129636	Peptide #3520 enco
74	6	3.4	34	22	AA104838	Human peptide enco
75	6	3.4	34	23	AA138931	Peptide #7970 enco
76	6	3.4	37	22	AA124809	Human brain expres
77	6	3.4	37	22	AA161286	Human bone marrow
78	6	3.4	37	22	AA174012	Peptide #6628 enco
79	6	3.4	37	22	AA120194	Peptide #8221 enco
80	6	3.4	37	22	AA134184	Human peptide enco
81	6	3.4	37	23	AA143907	Human neurofilamen
82	6	3.4	37	23	AA143907	Human neurofilamen
83	6	3.4	44	19	AA120741	Human neurofilamen

84	6	3.4	49	21	AA656775	Arabidopsis, thalia	157	6	3.4	96	22	AAU67835	Propionibacterium
85	6	3.4	50	23	ABP06117	Human ORFX protein	158	6	3.4	98	22	AAU62250	Propionibacterium
86	6	3.4	53	20	AA031462	A. thaliana RAP2.7	159	6	3.4	98	22	AAW91797	Human immune/haema
87	6	3.4	53	21	AA025821	RAP2.7-R2 direct r	160	6	3.4	99	21	AB040876	Human ORFX ORF640
88	6	3.4	54	23	AA023550	Human FAIL Ig doma	161	6	3.4	99	23	AB060513	Human ORFX protein
89	6	3.4	55	20	AA123333	Human 5' EST secre	162	6	3.4	100	22	ABG00419	Novel human diagno
90	6	3.4	55	21	AA030328	Arabidopsis thalia	163	6	3.4	100	22	AAU31707	Novel human diagno
91	6	3.4	55	21	AA030328	Human secreted pro	164	6	3.4	101	21	AA000797	Human secreted pro
92	6	3.4	58	22	AAU56725	Propionibacterium	165	6	3.4	104	22	ABG29606	Novel human diagno
93	6	3.4	63	21	AA055635	Arabidopsis thalia	166	6	3.4	104	22	ABG43053	Peptide #10559 enc
94	6	3.4	63	21	AA058828	Arabidopsis thalia	167	6	3.4	104	22	ABG26224	Protein #8223 enco
95	6	3.4	64	22	AAU58118	Propionibacterium	168	6	3.4	104	22	AAW63955	Human brain expres
96	6	3.4	64	22	ABG28986	Peptide #1637 enco	169	6	3.4	104	22	AAW76775	Human bone marrow
97	6	3.4	64	22	ABG34154	Peptide #1637 enco	170	6	3.4	104	22	AAW20996	Peptide #7430 enco
98	6	3.4	64	22	ABG19598	Protein #1597 enco	171	6	3.4	104	22	AAW36882	Peptide #10919 enc
99	6	3.4	64	22	AAW54946	Human brain expres	172	6	3.4	104	22	ABG45940	Human peptide enco
100	6	3.4	64	22	AAW67326	Human bone marrow	173	6	3.4	105	22	ABG02829	Novel human diagno
101	6	3.4	64	22	AAW15161	Peptide #1595 enco	174	6	3.4	105	22	AAO13822	Human polypeptide
102	6	3.4	64	22	AAW27618	Peptide #1655 enco	175	6	3.4	106	22	AAO02661	Human polypeptide
103	6	3.4	64	22	AAW02899	Peptide #1581 enco	176	6	3.4	107	23	ABW97814	Human secretory po
104	6	3.4	64	23	ABG36969	Human peptide enco	177	6	3.4	109	21	AAV75634	Neisseria gonorrhoe
105	6	3.4	66	22	AAW89590	Human immune/haema	178	6	3.4	109	22	ABG28332	Novel human diagno
106	6	3.4	68	20	AAV31454	A. thaliana AP2 do	179	6	3.4	110	22	AAU86779	Novel human connec
107	6	3.4	68	21	AA025814	AP2 direct repeat	180	6	3.4	112	11	AAW05219	GX3262 antigen enc
108	6	3.4	68	22	ABG43732	Peptide #11238 enc	181	6	3.4	112	22	AAU50091	Propionibacterium
109	6	3.4	68	22	ABG43924	Peptide #11430 enc	182	6	3.4	113	21	AAW03930	Human secreted pro
110	6	3.4	68	22	ABG16899	Human nervous syst	183	6	3.4	114	20	AAW97370	Human 1-40 OH grow
111	6	3.4	68	22	ABG26678	Protein #8677 enco	184	6	3.4	116	18	AAW28116	Amino acid sequenc
112	6	3.4	68	22	AAW64694	Human brain expres	185	6	3.4	119	19	AAW98819	H. pylori GHPO 141
113	6	3.4	68	22	AAW64914	Human brain expres	186	6	3.4	122	21	ABG33140	Eucalyptus grandis
114	6	3.4	68	22	AAW77469	Human bone marrow	187	6	3.4	122	22	ABW17626	Human nervous syst
115	6	3.4	68	22	AAW77650	Human bone marrow	188	6	3.4	122	22	AAO04820	Human polypeptide
116	6	3.4	68	22	AAW21395	Peptide #7829 enco	189	6	3.4	122	22	AAW76439	Human colon cancer
117	6	3.4	68	22	AAW21571	Peptide #8005 enco	190	6	3.4	123	21	AAW03927	Human secreted pro
118	6	3.4	68	22	AAW37630	Peptide #11667 enc	191	6	3.4	124	22	AAW05396	Human polypeptide
119	6	3.4	68	22	AAW37850	Peptide #11887 enc	192	6	3.4	125	20	AAW97365	Bovine growth horm
120	6	3.4	68	23	ABG46487	Human peptide enco	193	6	3.4	125	20	AAW97366	Porcine growth hor
121	6	3.4	68	23	ABG46682	Human peptide enco	194	6	3.4	125	20	AAW97367	Ovine growth hormo
122	6	3.4	69	20	AAV31467	N. tabacum EREBP-1	195	6	3.4	125	20	AAW97369	Caprine growth hor
123	6	3.4	69	21	AAW56970	Human prostate can	196	6	3.4	126	18	AAW27648	Secreted protein B
124	6	3.4	69	21	AAW25826	AP2 domain amino a	197	6	3.4	126	18	AAW44084	Human secreted pro
125	6	3.4	69	22	AAW89883	Human immune/haema	198	6	3.4	126	21	AAW10240	Human adult ovary
126	6	3.4	69	22	AAW73849	Human colon cancer	199	6	3.4	131	20	AAW73843	Human prostate tum
127	6	3.4	69	23	ABP10635	Human ORFX protein	200	6	3.4	131	22	ABG13883	Novel human diagno
128	6	3.4	71	21	AAU03201	Human secreted pro	201	6	3.4	132	22	ABG27033	Novel human diagno
129	6	3.4	72	22	AAU03921	Propionibacterium	202	6	3.4	132	22	AAW90564	C glutamicum prote
130	6	3.4	72	23	ABP08259	Human ORFX protein	203	6	3.4	132	23	ABG60172	Human DTHP polype
131	6	3.4	76	22	AAU18559	Human lung antigen	204	6	3.4	133	22	AAW28234	S. epidermidis Ope
132	6	3.4	77	23	ABP33821	Human hydrolase-11	205	6	3.4	135	22	AAO04104	Human polypeptide
133	6	3.4	77	23	ABG55088	Lactococcus lactis	206	6	3.4	136	18	AAW24536	Immunoglobulin RB6
134	6	3.4	78	23	ABP26123	Streptococcus poly	207	6	3.4	136	22	ABG28271	Novel human diagno
135	6	3.4	78	23	ABP26124	Streptococcus poly	208	6	3.4	137	22	ABW09065	Thermus caldophilu
136	6	3.4	79	23	ABP02069	Human ORFX protein	209	6	3.4	137	23	AAW05552	Human G protein-co
137	6	3.4	80	23	AAU83080	Novel secreted pro	210	6	3.4	141	21	AAW40829	Human ORFX ORF593
138	6	3.4	82	20	AAV76592	Human ovarian tumo	211	6	3.4	141	23	ABF03564	Human ORFX protein
139	6	3.4	82	22	AAW24866	C glutamicum prote	212	6	3.4	142	21	ABF59091	Breast and ovarian
140	6	3.4	82	23	ABP27759	Streptococcus poly	213	6	3.4	142	22	ABG27260	Novel human diagno
141	6	3.4	85	22	AAU46894	Propionibacterium	214	6	3.4	142	23	ABG60149	Human DTHP polype
142	6	3.4	85	22	AAU11082	Human polypeptide	215	6	3.4	145	21	AAW61604	Arabidopsis thalia
143	6	3.4	88	22	ABW17025	Human nervous syst	216	6	3.4	147	18	AAW28324	Staphylococcus aur
144	6	3.4	90	21	AAV75956	Rat skin cell prot	217	6	3.4	149	20	AAW02696	Human secreted pro
145	6	3.4	90	22	ABG26441	Novel human diagno	218	6	3.4	152	21	AAW05317	Arabidopsis thalia
146	6	3.4	90	22	ABG38706	Peptide #6212 enco	219	6	3.4	153	21	AAW57110	Human prostate can
147	6	3.4	90	22	AAW59339	Human brain expres	220	6	3.4	155	21	AAW33141	Pinus radiata tran
148	6	3.4	90	22	AAW71892	Human bone marrow	221	6	3.4	155	22	ABG10871	Novel human diagno
149	6	3.4	90	22	AAW32172	Peptide #6209 enco	222	6	3.4	156	22	ABG04613	Novel human diagno
150	6	3.4	90	22	AAW55895	Skin cell protein,	223	6	3.4	159	22	ABG20837	Novel human diagno
151	6	3.4	90	23	ABG41705	Human peptide enco	224	6	3.4	161	22	ABG12607	Novel human diagno
152	6	3.4	90	23	ABW72095	Rat protein isolat	225	6	3.4	161	22	AAW09519	Human polypeptide
153	6	3.4	91	22	AAW57236	Human brain expres	226	6	3.4	162	18	AAW89772	Staphylococcus aur
154	6	3.4	91	22	AAW29971	Peptide #4008 enco	227	6	3.4	162	21	AAW08128	A polyphenol oxida
155	6	3.4	94	21	AAW16277	Arabidopsis thalia	228	6	3.4	168	19	AAW75107	Human secreted pro
156	6	3.4	95	21	AAW34533	Human secreted pro	229	6	3.4	170	21	AAW09728	Arabidopsis thalia

230	6	3.4	174	21	AAB41911	Human ORFX ORF1675	303	6	3.4	267	23	AAO15079	Human GABABR1b tra
231	6	3.4	175	15	AAR62783	Borrelia VS461 ant	304	6	3.4	268	23	ABP29308	Streptococcus poly
232	6	3.4	176	15	AAR62788	Borrelia 20047 ant	305	6	3.4	269	19	ABY85990	S. pneumoniae derl
233	6	3.4	178	15	AAR62776	Borrelia SIMON ant	306	6	3.4	270	22	ABB69971	Drosophila melanog
234	6	3.4	178	15	AAR60890	Borrelia SIMON ant	307	6	3.4	273	21	ABG36285	Arabidopsis thalia
235	6	3.4	181	23	ABB90622	Chlamydia pneumonia	308	6	3.4	273	21	AAG40304	Arabidopsis thalia
236	6	3.4	181	23	ABB90677	Chlamydia pneumonia	309	6	3.4	276	19	AAW61204	Streptococcus sp07
237	6	3.4	184	21	AAB24445	Human secreted pro	310	6	3.4	276	23	ABP54623	S. pneumoniae pnu
238	6	3.4	185	20	AAV35791	Amino acid sequenc	311	6	3.4	277	21	AAB13593	Streptomyces globi
239	6	3.4	189	15	AAR60909	Borrelia H13 antig	312	6	3.4	277	23	ABB89958	Human polypeptide
240	6	3.4	191	15	AAR60897	Borrelia VS461 ant	313	6	3.4	278	12	AAR13338	Mac-2 protein incl
241	6	3.4	191	15	AAR60902	Borrelia VS461 ant	314	6	3.4	278	21	AAAB56745	Human prostate can
242	6	3.4	191	22	AAU87130	Novel central nerv	315	6	3.4	278	23	AAE23547	Human FAII extrace
243	6	3.4	191	22	AAU86766	Novel human connec	316	6	3.4	280	22	ABB71298	Drosophila melanog
244	6	3.4	191	22	ABG26443	Drosophila melanog	317	6	3.4	282	20	AAV14483	Fragment of human
245	6	3.4	192	22	ABG26443	Novel human diagno	318	6	3.4	288	21	AAAG57319	Arabidopsis thalia
246	6	3.4	192	23	ABP00179	Human ORFX protein	319	6	3.4	288	23	ABP30012	Streptococcus poly
247	6	3.4	194	22	AAU69472	Human purified sec	320	6	3.4	292	23	ABP30012	Streptococcus poly
248	6	3.4	196	22	ABG11745	Novel human diagno	321	6	3.4	296	17	AAW11582	Human arginine met
249	6	3.4	207	22	ABG20139	Novel human diagno	322	6	3.4	296	21	AAAG13941	Streptomyces prist
250	6	3.4	214	20	AAV06303	Mouse pancreatic-d	323	6	3.4	296	21	ABG37305	Arabidopsis thalia
251	6	3.4	216	10	AAV93703	Sequence of the 20	324	6	3.4	296	23	ABB92656	Arabidopsis thalia
252	6	3.4	216	18	AAW31582	Elmeria tenella sp	325	6	3.4	298	21	AAAG09727	Herbicidically activ
253	6	3.4	217	22	ABB10356	Human cDNA SEQ ID	326	6	3.4	298	16	AAAG66405	Arabidopsis thalia
254	6	3.4	220	21	ABAB33001	Pinus radiata tran	327	6	3.4	299	21	AAAG09726	Bg/1 modification
255	6	3.4	222	22	ABG16504	Novel human diagno	328	6	3.4	299	21	AAAG09726	Arabidopsis thalia
256	6	3.4	223	20	AAV36346	Fragment of human	329	6	3.4	302	21	AAAG04966	Arabidopsis thalia
257	6	3.4	223	22	ABB12247	Human PI 5-phospha	330	6	3.4	304	21	AAAG37304	Arabidopsis thalia
258	6	3.4	226	22	AAAG39182	Human polypeptide	331	6	3.4	305	21	AAAB58916	Breast and ovarian
259	6	3.4	228	18	AAW20224	H. pylori transpor	332	6	3.4	305	23	AAE23557	Arabidopsis thalia
260	6	3.4	230	22	ABB70669	Drosophila melanog	333	6	3.4	307	22	ABB51275	Human FAII protein
261	6	3.4	231	21	AAAG13942	Arabidopsis thalia	334	6	3.4	307	22	ABB51275	Human secreted pro
262	6	3.4	231	21	AAAG37306	Arabidopsis thalia	335	6	3.4	308	21	AAAG21127	Arabidopsis thalia
263	6	3.4	233	22	AAAG33251	Human protein HP02	336	6	3.4	310	23	ABP43485	Human secreted pro
264	6	3.4	234	19	AAW37089	Lycopodium escul	337	6	3.4	311	21	AAAG46265	Arabidopsis thalia
265	6	3.4	235	21	AAAG29466	Arabidopsis thalia	338	6	3.4	314	20	AAV35552	Chlamydia pneumoni
266	6	3.4	236	22	ABG58682	Drosophila melanog	339	6	3.4	315	22	AAV35552	Human polypeptide
267	6	3.4	237	10	AAV90117	Group B Immunogen	340	6	3.4	315	22	AAAG39424	Human polypeptide
268	6	3.4	237	10	AAV90117	Group B Immunogen	341	6	3.4	318	21	AAAG36284	Arabidopsis thalia
269	6	3.4	237	19	AAW26000	Group B Elmeria te	342	6	3.4	318	21	AAAG40303	Arabidopsis thalia
270	6	3.4	239	23	ABP29167	Streptococcus poly	343	6	3.4	321	22	ABB51277	Human secreted pro
271	6	3.4	240	18	AAW25784	Anti-gp130 antibod	344	6	3.4	321	23	ABG60935	Novel floral meris
272	6	3.4	240	18	AAW20555	Helicobacter pylor	345	6	3.4	323	21	AAAG07631	Arabidopsis thalia
273	6	3.4	240	18	AAW24687	H. pylori cytoplas	346	6	3.4	323	21	AAAG18224	Arabidopsis thalia
274	6	3.4	241	18	AAW20668	H. pylori transpor	347	6	3.4	323	21	AAAG36807	Arabidopsis thalia
275	6	3.4	243	22	ABG13934	Novel human diagno	348	6	3.4	323	23	ABB93042	Herbicidically activ
276	6	3.4	247	21	AAAB12848	Human V-type ATPas	349	6	3.4	324	21	AAAG07630	Arabidopsis thalia
277	6	3.4	247	22	AAU32456	Novel human secret	350	6	3.4	324	21	ABG36806	Arabidopsis thalia
278	6	3.4	248	11	AAV05221	Antigen GX32621Ex	351	6	3.4	324	22	ABG07010	Novel human diagno
279	6	3.4	250	21	AAAB42488	Human ORFX ORF252	352	6	3.4	325	21	AAAB19393	Amino acid sequenc
280	6	3.4	250	21	AAAG18225	Arabidopsis thalia	353	6	3.4	325	21	AAAG05322	Arabidopsis thalia
281	6	3.4	252	22	AAW40968	Human polypeptide	354	6	3.4	325	21	AAAG06659	Arabidopsis thalia
282	6	3.4	252	22	AAW79289	Human protein SEQ	355	6	3.4	325	21	AAAG08960	Arabidopsis thalia
283	6	3.4	253	21	AAAG36286	Arabidopsis thalia	356	6	3.4	325	21	AAAG49308	Arabidopsis thalia
284	6	3.4	253	21	AAAG40305	Arabidopsis thalia	357	6	3.4	325	21	AAAG57818	Arabidopsis thalia
285	6	3.4	253	22	ABB64223	Drosophila melanog	358	6	3.4	325	22	AAU71864	C. glutamicum cyst
286	6	3.4	254	21	AAAG21129	Arabidopsis thalia	359	6	3.4	326	21	AAAG22881	Human peroxisome a
287	6	3.4	258	19	AAW61954	Rat galectin amino	360	6	3.4	326	21	AAAG46264	Arabidopsis thalia
288	6	3.4	259	20	AAV34948	Chlamydia pneumonia	361	6	3.4	326	22	AAW41210	Human polypeptide
289	6	3.4	259	21	AAAG21128	Arabidopsis thalia	362	6	3.4	327	21	AAAG18223	Arabidopsis thalia
290	6	3.4	261	22	ABG60232	Drosophila melanog	363	6	3.4	327	21	AAAG57318	Arabidopsis thalia
291	6	3.4	262	14	AAW42200	IgE binding protein	364	6	3.4	328	20	AAV74117	Human prostate tum
292	6	3.4	262	22	AAU25441	Human mdtt protein	365	6	3.4	328	21	AAAG44530	Arabidopsis thalia
293	6	3.4	262	23	AAU97821	Rat cell membrane	366	6	3.4	329	23	AAW50863	Arabidopsis thalia
294	6	3.4	263	23	AAU97822	Mouse cell membran	367	6	3.4	330	21	AAAG07629	Arabidopsis thalia
295	6	3.4	264	12	AAW12531	Mac2.16 expression	368	6	3.4	330	21	AAAG36805	Arabidopsis thalia
296	6	3.4	264	19	AAW71219	Beta-D-galactoside	369	6	3.4	330	22	AAU57276	Propionibacterium
297	6	3.4	265	22	ABW8462	Human membrane or	370	6	3.4	332	22	AAW49234	Mouse SNORF33 rece
298	6	3.4	267	15	AAW56259	K. marxianus URA3.	371	6	3.4	333	21	AAAG44632	A. thaliana putati
299	6	3.4	267	20	AAV49111	Human GABABR1a tra	372	6	3.4	333	22	AAAB5457	Arabidopsis AIR sy
300	6	3.4	267	20	AAV49112	Human GABABR1b tra	373	6	3.4	336	21	AAAG46263	Arabidopsis thalia
301	6	3.4	267	22	AAW93536	Human polypeptide,	374	6	3.4	340	21	AAAG04965	Arabidopsis thalia
302	6	3.4	267	23	AAO15078	Human GABABR1a tra	375	6	3.4	340	21	AAAG08959	Arabidopsis thalia

376	6	3.4	340	21	AAG57817	Arabidopsis thalia	449	6	3.4	407	22	ABG21343	Novel human diagno
377	6	3.4	340	22	AAM01012	CFE II protein seq	450	6	3.4	407	22	ABG10439	Novel human diagno
378	6	3.4	340	23	AAE22586	BS203 consensus pr	451	6	3.4	410	20	AAI37660	Chlamydia trachoma
379	6	3.4	341	19	AAW58573	Trichoderma reesei	452	6	3.4	410	22	AAB04102	Botulism toxin hea
380	6	3.4	341	21	AAG57816	Arabidopsis thalia	453	6	3.4	411	21	AAT76047	Rat skin cell prot
381	6	3.4	342	23	AAE22916	Human transporter	454	6	3.4	411	21	AAW55986	Skin cell protein,
382	6	3.4	343	23	AAE23546	Human mature FAIL	455	6	3.4	411	23	ABB72186	Rat protein isolat
383	6	3.4	346	21	AAG36678	Arabidopsis thalia	456	6	3.4	413	21	AAG04964	Arabidopsis thalia
384	6	3.4	346	23	ABB91294	Herbicidally activ	457	6	3.4	414	21	AAG15375	Arabidopsis thalia
385	6	3.4	349	22	ABG16891	Novel human diagno	458	6	3.4	415	22	AAM40868	Human polypeptide
386	6	3.4	349	22	AAG67525	Amino acid sequenc	459	6	3.4	415	22	AAM40869	Human polypeptide
387	6	3.4	350	21	AAG36677	Arabidopsis thalia	460	6	3.4	416	20	AAW88745	Human secreted pro
388	6	3.4	350	23	ABBS3286	Human polypeptide	461	6	3.4	416	22	ABB50513	Human secreted pro
389	6	3.4	352	21	AAG05321	Arabidopsis thalia	462	6	3.4	416	22	AAW25861	Human protein sequ
390	6	3.4	352	21	AAG49307	Arabidopsis thalia	463	6	3.4	417	21	AAB41474	Human ORFX ORF1238
391	6	3.4	353	16	AAW53388	Japanese cedar pol	464	6	3.4	418	22	AAB88368	Human membrane or
392	6	3.4	353	17	AAW81587	Cedar pollen aller	465	6	3.4	419	22	AAB04095	Botulism toxin hea
393	6	3.4	358	21	AAW94336	Human cell surface	466	6	3.4	421	22	AAG63222	Amino acid sequenc
394	6	3.4	359	20	AAW41690	Human PRO329 prote	467	6	3.4	426	23	AAE23289	Human nectin-3alph
395	6	3.4	359	21	AAB44246	Human PRO329 (UNQ2	468	6	3.4	432	21	AAG29464	Arabidopsis thalia
396	6	3.4	359	21	AAB34744	Human secreted pro	469	6	3.4	433	21	AAG15374	Arabidopsis thalia
397	6	3.4	359	22	AAU29044	Human PRO polypept	470	6	3.4	434	22	AAU36234	Pseudomonas aerugi
398	6	3.4	359	22	AAE03451	Human gene 25 enco	471	6	3.4	435	21	AAG14364	Arabidopsis thalia
399	6	3.4	359	23	ABG63386	Human albumin fusi	472	6	3.4	435	22	ABB11390	Human secreted pro
400	6	3.4	359	23	ABP39997	Staphylococcus epi	473	6	3.4	435	22	AAW80099	Human secreted pro
401	6	3.4	362	21	AAG06658	Arabidopsis thalia	474	6	3.4	436	22	ABG18343	Novel human diagno
402	6	3.4	362	21	AAG36676	Arabidopsis thalia	475	6	3.4	437	23	AAE23299	Amino acid sequenc
403	6	3.4	363	20	AAW37863	Amino acid sequenc	476	6	3.4	438	22	AAG63984	Mouse nectin-3gamm
404	6	3.4	363	20	AAW36333	Human secreted pro	477	6	3.4	442	22	AAG82974	S. epidermidis ope
405	6	3.4	363	22	AAW88563	Secreted protein e	478	6	3.4	442	22	AAE23293	Streptomyces galli
406	6	3.4	363	22	ABBS0330	Human secreted pro	479	6	3.4	443	22	AAE00483	Arabidopsis thalia
407	6	3.4	363	22	ABB51276	Human secreted pro	480	6	3.4	447	21	AAG15496	Human protein sequ
408	6	3.4	364	22	AAG29465	Arabidopsis thalia	481	6	3.4	447	22	AAB94438	Human uptake ABC tr
409	6	3.4	364	22	ABG14391	Novel human diagno	482	6	3.4	448	23	ABG61513	Arabidopsis th

522	6	3.4	497	22	AAG82265	S. epidermidis ope	595	6	3.4	560	23	AAU77628	S. agalactiae Sbp2
523	6	3.4	498	22	ABG23519	Novel human diagno	596	6	3.4	563	20	AAV14111	Human GABAB recept
524	6	3.4	501	23	ABP25949	Streptococcus poly	597	6	3.4	563	22	ABG72358	Virulent group B S
525	6	3.4	502	22	AAU33945	Staphylococcus aur	598	6	3.4	565	22	ABG18709	Novel human diagno
526	6	3.4	502	22	AAU36794	Staphylococcus aur	599	6	3.4	569	23	ABG92340	Herbicidally activ
527	6	3.4	502	22	AAU37228	Staphylococcus aur	600	6	3.4	571	21	AAG31041	Arabidopsis thalia
528	6	3.4	502	22	AAW79303	Human protein SEQ	601	6	3.4	574	23	ABG97289	Novel human protei
529	6	3.4	503	22	AAG82358	S. epidermidis ope	602	6	3.4	575	15	AAR54220	L.lactis alpha-ace
530	6	3.4	503	23	ABP39369	Staphylococcus epi	603	6	3.4	577	22	AAW78367	Human protein SEQ
531	6	3.4	504	22	ABG52616	Escherichia coli p	604	6	3.4	577	22	AAE06804	Mature human neuro
532	6	3.4	504	23	AAE23284	Human deleted nect	605	6	3.4	580	22	ABG18705	Novel human diagno
533	6	3.4	507	21	AAG36462	Arabidopsis thalia	606	6	3.4	581	23	ABG93540	Herbicidally activ
534	6	3.4	507	21	AAV81580	Streptococcus pneu	607	6	3.4	582	22	ABG06711	Novel human diagno
535	6	3.4	507	23	ABP40226	Staphylococcus epi	608	6	3.4	583	22	ABG16197	Novel human diagno
536	6	3.4	508	21	AAG14362	Arabidopsis thalia	609	6	3.4	584	22	AAW00930	Human bone marrow
537	6	3.4	510	22	AAG63983	Amino acid sequenc	610	6	3.4	585	22	AAW79351	Human protein SEQ
538	6	3.4	510	23	AAE23285	Mouse nectin-3-hum	611	6	3.4	586	22	ABG13054	Novel human diagno
539	6	3.4	510	23	AAE23286	Human nectin-3beta	612	6	3.4	588	17	AAW05297	Esterase secretory
540	6	3.4	510	23	AAE23292	Mouse nectin-3beta	613	6	3.4	594	19	AAW74580	Human membrane pro
541	6	3.4	510	23	AAU75483	S. aureus antigeni	614	6	3.4	594	22	ABG26986	Novel human diagno
542	6	3.4	511	21	AAG58809	Arabidopsis thalia	615	6	3.4	594	22	AAW94068	Human protein sequ
543	6	3.4	511	19	AAW68506	Human acid sensng	616	6	3.4	595	21	AAG58807	Arabidopsis thalia
544	6	3.4	512	20	AAW93420	Human BNC1 protein	617	6	3.4	595	23	AAE23288	Human nectin-3beta
545	6	3.4	512	21	AAV69177	A human acid-sensi	618	6	3.4	597	21	AAW08122	A polyphenol oxida
546	6	3.4	512	23	AAU10903	Human acid sensng	619	6	3.4	606	22	AAE06799	Human neuronal gui
547	6	3.4	512	23	AAU10905	Rat acid sensng i	620	6	3.4	606	23	ABP60969	Novel human protei
548	6	3.4	515	22	AAW80287	Human protein SEQ	621	6	3.4	606	23	ABP60970	Novel human protei
549	6	3.4	516	22	AAW81891	Novel human diagno	622	6	3.4	606	23	AAU96849	Human NOV1a protei
550	6	3.4	517	23	ABP41923	Human ovarian anti	623	6	3.4	606	23	AAU96850	Human NOV4b protei
551	6	3.4	518	23	ABG54962	Lactococcus lactis	624	6	3.4	606	23	ABG60229	Human NOV4b protei
552	6	3.4	520	13	AAK21075	Sequence encoded b	625	6	3.4	606	23	ABG60230	Human Slit-3-like
553	6	3.4	520	13	AAK21076	Sequence encoded b	626	6	3.4	606	23	ABG60230	Human Slit-3-like
554	6	3.4	520	13	AAK21077	Sequence encoded b	627	6	3.4	606	23	AAU85410	Human protein NOV8
555	6	3.4	521	21	AAK13578	Streptomyces globi	628	6	3.4	606	23	AAU85411	Human protein NOV9
556	6	3.4	521	21	AAK13605	Streptomyces globi	629	6	3.4	611	22	AAW85021	NOV8 allelic varia
557	6	3.4	521	21	AAK13605	Streptomyces globi	630	6	3.4	612	21	AAW85021	Shrimp white spot
558	6	3.4	524	21	AAK13605	Human ORFX ORF1664	631	6	3.4	612	21	AAW85021	Arabidopsis thalia
559	6	3.4	524	21	AAK13605	Human transmembran	632	6	3.4	613	22	AAW85021	Human polypeptide
560	6	3.4	524	23	ABP39470	Staphylococcus epi	633	6	3.4	614	19	AAW83830	Specific co-activa
561	6	3.4	529	21	AAK32200	Human receptor mol	634	6	3.4	614	19	AAW83830	Amino acid sequenc
562	6	3.4	533	19	AAW59442	Hordeum vulgare ML	635	6	3.4	620	20	AAV13357	Amino acid sequenc
563	6	3.4	533	19	AAW59443	Hordeum vulgare ML	636	6	3.4	620	20	AAV13357	Human PRO227 poly
564	6	3.4	533	21	AAW3401	Barley Mlo protein	637	6	3.4	620	22	AAW74705	Human membrane ass
565	6	3.4	534	20	AAW26966	Wheat Mlo fungal r	638	6	3.4	621	19	AAW80225	Human PRO227 prote
566	6	3.4	534	20	AAW26967	Wheat Mlo fungal r	639	6	3.4	621	19	AAW80225	Helicobacter pylor
567	6	3.4	534	21	AAW26968	Wheat Mlo fungal r	640	6	3.4	621	21	AAW11121	Human GAMT-1 prote
568	6	3.4	534	21	AAW26968	Wheat Mlo homologo	641	6	3.4	621	21	AAW57559	Human growth-assoc
569	6	3.4	534	22	AAW26968	Amino acid sequenc	642	6	3.4	621	22	AAW35710	Helicobacter pylor
570	6	3.4	534	22	AAW26968	Amino acid sequenc	643	6	3.4	621	22	AAW35710	Helicobacter pylor
571	6	3.4	534	22	AAW26968	Amino acid sequenc	644	6	3.4	622	10	AAW92618	Human protein sequ
572	6	3.4	537	23	ABP28881	Streptococcus poly	645	6	3.4	622	10	AAW92618	Rhoptry membrane a
573	6	3.4	542	23	ABW54556	Lactococcus lactis	646	6	3.4	622	13	AAW27532	Plasmodium falci
574	6	3.4	542	23	ABW54556	Lactococcus lactis	647	6	3.4	622	16	AAW68840	Thermococcus amida
575	6	3.4	544	17	AAW03943	Human deleted nect	648	6	3.4	622	19	AAW41248	Thermococcus amida
576	6	3.4	544	19	AAW79570	LKT-GaRH protein f	649	6	3.4	622	22	AAW93427	Human polypeptide,
577	6	3.4	544	23	ABP27253	LKT-GaRH chimeric	650	6	3.4	622	22	AAW88452	Human membrane or
578	6	3.4	545	22	AAW41501	Streptococcus poly	651	6	3.4	622	23	AAW99359	Synthetic P. falci
579	6	3.4	546	21	AAV51773	Human polypeptide	652	6	3.4	622	23	AAW99359	P. falci parum apic
580	6	3.4	546	21	AAV51773	A. pleuropneumonia	653	6	3.4	622	23	ABP43477	Human secreted pro
581	6	3.4	547	14	AAW34668	A. pleuropneumonia	654	6	3.4	624	21	AAW31314	Arabidopsis thalia
582	6	3.4	547	22	AAW34668	Ap Serotype 7 60KD	655	6	3.4	626	22	AAW94393	Human protein sequ
583	6	3.4	548	22	AAW34668	Human AFP protein	656	6	3.4	627	21	AAW31989	Arabidopsis thalia
584	6	3.4	548	22	AAW34668	Amino acid sequenc	657	6	3.4	632	23	AAE23287	Arabidopsis thalia
585	6	3.4	549	22	AAW34668	Amino acid sequenc	658	6	3.4	634	23	AAE23287	Human nectin-3alp
586	6	3.4	549	23	AAE23282	Mouse nectin-3-hum	659	6	3.4	638	21	AAV75343	Neisseria gonorrhe
587	6	3.4	549	23	AAE23282	Human nectin-3alp	660	6	3.4	642	21	AAW43122	Human ORFX ORF2886
588	6	3.4	550	21	AAW58808	Mouse nectin-3alp	661	6	3.4	644	20	AAV35717	C. pneumoniae prot
589	6	3.4	550	22	AAW58808	Arabidopsis thalia	662	6	3.4	650	21	AAW31502	Arabidopsis thalia
590	6	3.4	550	22	AAW58808	Human polypeptide	663	6	3.4	650	21	AAW82527	Human RP105 protei
591	6	3.4	559	22	ABG63943	Drosophila melanog	664	6	3.4	652	21	AAW31501	Arabidopsis thalia
592	6	3.4	559	22	AAW39715	Human polypeptide	665	6	3.4	654	22	AAW42042	Human polypeptide
593	6	3.4	559	22	AAW39715	Human polypeptide	666	6	3.4	658	22	AAU35354	Enterococcus faeca
594	6	3.4	559	22	AAW39715	Human polypeptide	667	6	3.4	661	18	AAW28510	Product of clone J
						Human secreted pro	668	6	3.4	661	19	AAW47274	Human B-cell activ
						Human ORFX ORF2802	669	6	3.4	661	20	AAW87556	B cell surface pro
							670	6	3.4	662	16	AAW73595	Cotransporter prot



814	6	3.4	960	20	AAV28841	Rat GABAB1a recep	887	1375	23	AAU79288	Streptococcus muta
815	6	3.4	960	20	AAV49122	Rat GABAB1a prote	888	1382	23	ABB76484	Human prostate-spe
816	6	3.4	960	20	AAV29797	Murine gamma-amino	889	1383	18	AAW33227	Procoagulant-activ
817	6	3.4	960	21	AAV83145	Human GABAB1Aa rec	890	1383	18	AAW33228	Procoagulant-activ
818	6	3.4	960	22	ABW50089	Murine GABA-B-R1a	891	1383	18	AAW33229	Procoagulant-activ
819	6	3.4	960	23	AAO15089	Murine GABA-B-R1a	892	1403	19	AAW79574	Leukotoxin 101. S
820	6	3.4	961	20	ABW50633	Murine GABA-B rece	893	1424	9	AAW80268	Modified factor VII
821	6	3.4	961	20	AAV28838	Human GABAB1a rec	894	1424	10	AAW91169	Sequence of 740 Ar
822	6	3.4	961	20	AAV29798	Human gamma-amino-	895	1424	22	AAW48842	Mutant mature huma
823	6	3.4	961	20	AAV14101	Human GABAB recep	896	1425	9	AAW80267	Modified factor VI
824	6	3.4	961	23	AAO14254	Human GABAB recep	897	1431	22	AAW67960	Amino acid sequenc
825	6	3.4	961	23	ABW50632	Human GABA-B recep	898	1438	21	AAW01262	B-domain deleted f
826	6	3.4	962	22	AAW50090	Human GABA-B-R1a	899	1440	12	AAW12971	Factor VIII:SQ
827	6	3.4	964	20	AAV14105	Canine GABAB recep	900	1443	19	AAW44137	Homo sapiens facto
828	6	3.4	964	23	AAV72538	Arabidopsis cell c	901	1443	20	AAW31598	Porcine factor VII
829	6	3.4	969	22	ABW63009	Drosophila melanog	902	1443	22	AAW50469	Porcine factor VII
830	6	3.4	975	22	ABW60615	Drosophila melanog	903	1457	19	AAW46246	Human factor VIII
831	6	3.4	975	22	ABW93230	Human protein sequ	904	1457	19	AAW44372	Human factor VIII
832	6	3.4	976	22	ABW63981	Drosophila melanog	905	1457	20	AAW21675	Beta-domain delete
833	6	3.4	977	17	AAW03942	LKT-GnRH protein f	906	1459	22	AAW10827	Human factor VIII
834	6	3.4	977	19	AAW79569	LKT-GnRH chimeric	907	1459	22	AAW10832	Human factor VIII
835	6	3.4	977	22	AAW39082	Human polypeptide	908	1459	22	AAW10833	Human factor VIII
836	6	3.4	992	16	AAW73917	Rubella virus Ther	909	1467	22	AAW11207	Modified porcine f
837	6	3.4	1006	22	AAW81530	S. epidermidis ope	910	1468	20	AAW38825	Neisseria gonorrhoe
838	6	3.4	1014	8	AAW71139	Factor VIII:c varia	911	1470	22	AAW31518	Amino acid sequenc
839	6	3.4	1017	23	AAU79285	Streptococcus muta	912	1471	18	AAW23414	Human B-domain del
840	6	3.4	1018	15	AAW63759	Human contactin (E	913	1471	22	AAW67959	Amino acid sequenc
841	6	3.4	1018	17	AAW87028	Human contactin.	914	1475	22	ABW21224	Novel human diagno
842	6	3.4	1061	20	AAW87504	Human N-methyl-D-a	915	1475	23	AAU98027	S. mutans glucosyl
843	6	3.4	1069	15	AAW52748	Bovine IFNgamma/LK	916	1475	23	AAU98030	S. mutans glucosyl
844	6	3.4	1069	18	AAW13867	Chimeric protein #	917	1475	23	AAU98031	S. mutans glucosyl
845	6	3.4	1069	21	AAW21074	Bovine gamma-IFN/P	918	1475	23	AAU98032	S. mutans glucosyl
846	6	3.4	1076	23	ABP39192	Staphylococcus epi	919	1475	23	AAU98033	S. mutans glucosyl
847	6	3.4	1081	21	AAW26240	Human N-methyl-D-a	920	1475	23	AAU98034	S. mutans glucosyl
848	6	3.4	1081	23	ABW56523	Human NMDA recepto	921	1475	23	AAU98035	S. mutans glucosyl
849	6	3.4	1096	9	AAW2507	Pullulanase protei	922	1475	23	AAU98036	S. mutans glucosyl
850	6	3.4	1098	13	AAW22103	Bovine IL-2 - LKT	923	1475	23	AAU98037	S. mutans glucosyl
851	6	3.4	1098	15	AAW52747	Bovine IL-2/LKT ch	924	1475	23	AAU98038	S. mutans GTFB mut
852	6	3.4	1098	18	AAW13866	Chimeric protein #	925	1475	23	AAU98039	S. mutans glucosyl
853	6	3.4	1098	21	AAW21073	Bovine IL-2/Pasteu	926	1475	23	AAU98040	S. mutans glucosyl
854	6	3.4	1112	22	ABW62695	Drosophila melanog	927	1476	23	AAU79284	Streptococcus muta
855	6	3.4	1129	22	ABW20477	Novel human diagno	928	1492	23	ABW61768	Novel alpha-2-macr
856	6	3.4	1147	21	AAW30843	Arabidopsis thalia	929	1496	21	AAW43629	Arabidopsis thalia
857	6	3.4	1164	22	AAU03555	Human protein kina	930	1508	22	AAW63548	A human alpha-2 ma
858	6	3.4	1164	23	AAE19153	Human kinase poly	931	1511	21	AAW46260	Arabidopsis thalia
859	6	3.4	1172	22	ABW66459	Drosophila melanog	932	1514	23	ABW90837	Herbicidially activ
860	6	3.4	1212	20	AAW87503	Human N-methyl-D-a	933	1516	9	AAW80285	Modified factor VI
861	6	3.4	1212	21	AAW46262	Arabidopsis thalia	934	1598	21	AAW43628	Arabidopsis thalia
862	6	3.4	1213	22	AAU33862	Staphylococcus aur	935	1610	22	AAW30568	A full length huma
863	6	3.4	1215	21	AAW30842	Arabidopsis thalia	936	1612	21	AAW43627	Arabidopsis thalia
864	6	3.4	1217	22	AAU36930	Staphylococcus aur	937	1618	22	AAW5504	Human protein kina
865	6	3.4	1220	22	AAW86160	D. melanogaster GA	938	1632	22	ABW62676	Drosophila melanog
866	6	3.4	1221	22	ABW62615	Drosophila melanog	939	1647	22	ABW10750	Novel human diagno
867	6	3.4	1221	22	AAU38942	Drosophila G-prote	940	1661	18	AAW18670	Factor VIII-dB695-
868	6	3.4	1231	22	ABW59738	Drosophila melanog	941	1665	23	AAW24151	Human kinase (PKIN
869	6	3.4	1232	21	AAW26239	Human N-methyl-D-a	942	1665	23	AAW15372	Human myosin light
870	6	3.4	1232	23	AAW47961	Human NMDA recepto	943	1922	22	ABW63631	Arabidopsis thalia
871	6	3.4	1243	22	ABW21221	Novel human diagno	944	2066	22	AAE10146	Streptomyces nous
872	6	3.4	1268	22	ABW67095	Drosophila melanog	945	2090	22	ABW58261	Drosophila melanog
873	6	3.4	1285	22	ABW16337	Novel human diagno	946	2098	17	AAW86863	Factor-VIII. Homo
874	6	3.4	1291	22	ABW71544	Drosophila melanog	947	2115	21	AAW57849	Pig Factor VIII pr
875	6	3.4	1305	22	ABW59497	Drosophila melanog	948	2133	19	AAW44133	Sus scrofa factor
876	6	3.4	1305	22	AAW38927	Drosophila G-prote	949	2133	20	AAW31597	Porcine factor VII
877	6	3.4	1305	22	AAW86161	D. melanogaster GA	950	2133	22	AAW11203	Porcine factor VII
878	6	3.4	1318	22	AAU28169	Novel human secret	951	2133	22	AAW50468	Porcine factor VII
879	6	3.4	1323	20	AAW49133	GABA-BR1a*Gqo5 fus	952	2280	22	ABW61650	Drosophila melanog
880	6	3.4	1323	23	AAO15099	Human GABA-BR1a*Gq	953	2304	21	AAW57848	Drosophila melanog
881	6	3.4	1334	12	AAW14481	Lk1A::lacZ fusion p	954	2319	19	AAW53485	Mouse Factor VIII
882	6	3.4	1334	15	AAW50290	Lk1A::lacZ fusion	955	2319	19	AAW44135	Murine factor VIII
883	6	3.4	1352	21	AAW46261	Arabidopsis thalia	956	2319	20	AAW31596	Mus musculus facto
884	6	3.4	1355	21	AAW30841	Arabidopsis thalia	957	2319	22	AAW11202	Murine factor VIII
885	6	3.4	1373	21	AAW42190	Human ORFX ORF1954	958	2319	22	AAW50467	Murine factor VIII
886	6	3.4	1375	23	AAU98028	S. mutans glucosyl	959	2332	8	AAW71726	Factor VIII:c varia





CC comprise a immunogenic or immunostimulatory molecule or the peptides  
 CC may be modified with lipids, or linked to synthetic PRP as synthetic  
 CC lipoglycopeptide conjugates to produce alternative vaccines.

XX Sequence 35 AA;

Query Match 5.6%; Score 10; DB 14; Length 35;

Best Local Similarity 100.0%; Pred. No. 0.016;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 ELSGFHQLTD 33

|||||

Db 2 ELSGFHQLTD 11

RESULT 3

AAR12455  
 ID AAR12455 standard; Protein; 36 AA.

XX AC AAR12455;

XX 26-JUL-1991 (first entry)

DE HTBPI-14(400-437) 6U H.Influenzae Type b OMP epitope.

XX bacterial meningitis; vaccine; P1 gene; T-cell antigen;

KW outer membrane protein.

XX Haemophilus influenzae.

XX WO9106652-A.

XX 16-MAY-1991.

XX 31-OCT-1990; 90WO-CA00374.

XX 31-OCT-1989; 89GB-0024473.

XX (CONN-) CONNAUGHT LAB LTD.

PA (UNIW ) WASHINGTON UNIV ST LOUIS.

XX Munson RS, Grass S, Chong P, Yang Y, Fahim R, McVerry P;  
 PI Klein M;

XX WPI; 1991-164201/22.

XX Outer membrane protein of Haemophilus influenzae type B - used as  
 PT vaccine against infections, esp. in infants and for diagnosis

XX Disclosure; Table 2; 33pp; English.

XX Synthetic peptide corresponding to amino acids 400 to 437 of the H.  
 CC influenzae strain 8358 (OMP subtype 6U) P1 protein. The N-terminal  
 CC Cys residue is additional and allows coupling to a carrier in one  
 CC specific orientation. The free peptide was used as an immunogen to  
 CC produce antibodies showing that the peptide comprises both T-helper  
 CC determinant and B-cell epitope(s).  
 CC See also AAR12446-R12454 and AAR12083.

XX Sequence 36 AA;

Query Match 5.6%; Score 10; DB 12; Length 36;

Best Local Similarity 100.0%; Pred. No. 0.017;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 ANLYGLNLNY 177

|||||

Db 25 ANLYGLNLNY 34

RESULT 4

AAR40066  
 ID AAR40066 standard; peptide; 36 AA.

XX. AAR40066;  
 AC 04-FEB-1994 (first entry)  
 DT  
 XX  
 DE Hib OMP P1 peptide HIBP1-14 (400-437).  
 KW Haemophilus influenzae; type b; Hib; outer membrane protein; P1; P2;  
 KW P6; vaccine; antibody; detection; lipoglycopeptide conjugate;  
 KW immunogen.  
 XX Synthetic.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1  
 FT /note= "May be absent"  
 FT Misc-difference 5..14  
 FT /note= "Residues are different to those found in P1  
 FT protein of H. influenzae strain 1H"

XX WO9315205-A.

XX 05-AUG-1993.

XX 03-FEB-1993; 93WO-CA00041.

XX 03-FEB-1992; 92GB-0002219.

XX (CONN-) CONNAUGHT LAB LTD.

XX Chong P, Kandil A, Klein MH, Sia C;

XX WPI; 1993-258681/32.

XX Synthetic Haemophilus influenzae conjugate vaccine - comprising  
 PT T-helper cell determinants and B-cell epitope(s) linked to  
 PT synthetic oligo:saccharide(s)

XX Table 1; Page 48; 99pp; English.

XX The sequences given in AAR40053-101 are peptide fragments derived from  
 CC the Haemophilus influenzae type b (Hib) outer membrane proteins P1,  
 CC P2 and P6. These peptides may be used in a vaccine against Hib  
 CC infection and antibodies against these peptides may be used in test  
 CC kits to detect H. influenzae in a sample. The vaccine may further  
 CC comprise an immunogenic or immunostimulatory molecule or the peptides  
 CC may be modified with lipids, or linked to synthetic PRP as synthetic  
 CC lipoglycopeptide conjugates to produce alternative vaccines.

XX Sequence 36 AA;

Query Match 5.6%; Score 10; DB 14; Length 36;

Best Local Similarity 100.0%; Pred. No. 0.017;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 ANLYGLNLNY 177

|||||

Db 25 ANLYGLNLNY 34

RESULT 5

AAR12454

ID AAR12454 standard; Protein; 39 AA.

XX AAR12454;

XX 26-JUL-1991 (first entry)

DE HTBPI-13(400-437) 1H H.Influenzae Type b OMP epitope.

XX bacterial meningitis; vaccine; P1 gene; T-cell antigen;

KW outer membrane protein.

XX

OS Haemophilus influenzae.  
 PN WO9106652-A.  
 XX  
 PD 16-MAY-1991.  
 XX  
 PF 31-OCT-1990; 90WO-CA00374.  
 XX  
 PR 31-OCT-1989; 89GB-0024473.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 PA (UNIW) WASHINGTON UNIV ST LOUIS.  
 XX  
 PI Munson RS, Grass S, Chong P, Yang Y, Fahim R, McVerry P;  
 PI Klein M;  
 XX  
 DR WPI; 1991-164201/22.  
 XX  
 XX Outer membrane protein of Haemophilus influenzae type B - used as  
 PT vaccine against infections, esp. in infants and for diagnosis  
 PT  
 XX Disclosure; Table 2; 33pp; English.  
 PS  
 CC Synthetic peptide corresponding to amino acids 400 to 437 of the H.  
 CC influenzae P1 protein. The N-terminal Cys residue is additional and  
 CC allows coupling to a carrier in one specific orientation. The  
 CC free peptide was used as an immunogen to produce antibodies showing  
 CC that the peptide comprises both T-helper determinant and major  
 CC immunodominant B-cell epitope.  
 CC See also AAR12446-R12453 and AAR12455 and AAQ12083.  
 XX  
 XX Sequence 39 AA;  
 SQ

Query Match 5.6%; Score 10; DB 12; Length 39;  
 Best Local Similarity 100.0%; Pred. No. 0.018;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 168 ANLYGLNLNY 177  
 DB |||||  
 DB 28 ANLYGLNLNY 37

RESULT 6  
 AAR40065  
 ID AAR40065 standard; peptide; 39 AA.  
 XX  
 AC AAR40065;  
 XX  
 DT 04-FEB-1994 (first entry)  
 XX  
 DE Hib OMP P1 peptide HIBP1-13 (400-437).  
 XX  
 KW Haemophilus influenzae; type b; Hib; outer membrane protein; P1; P2;  
 KW P6; vaccine; antibody; detection; lipoglycopeptide conjugate;  
 KW immunogen.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..15  
 FT /label= C-P1  
 FT Region 16..53  
 FT /label= CHIBP2  
 XX  
 PN WO9315205-A.  
 XX  
 PD 05-AUG-1993.  
 XX  
 PF 03-FEB-1993; 93WO-CA00041.  
 XX  
 PR 03-FEB-1993;  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 XX  
 PI Chong P, Kandil A, Klein MH, Sia C;  
 XX  
 DR WPI; 1993-258681/32.  
 XX  
 PT Synthetic Haemophilus influenzae conjugate vaccine - comprising  
 PT T-helper cell determinants and B-cell epitope(s) linked to  
 PT synthetic oligo:saccharide(s)  
 XX  
 PS Table 11; Page 59; 99pp; English.  
 XX  
 CC The sequences given in AAR40053-101 are peptide fragments derived from  
 CC the Haemophilus influenzae type b (Hib) outer membrane proteins P1,  
 CC P2 and P6. These peptides may be used in a vaccine against Hib  
 CC infection and antibodies against these peptides may be used in test  
 CC kits to detect H. influenzae in a sample. The vaccine may further  
 CC comprise an immunogenic or immunostimulatory molecule or the peptides  
 CC may be modified with lipids, or linked to synthetic PRP as synthetic  
 CC lipoglycopeptide conjugates to produce alternative vaccines.  
 XX  
 XX Sequence 39 AA;  
 SQ

Query Match 5.6%; Score 10; DB 14; Length 39;  
 Best Local Similarity 100.0%; Pred. No. 0.018;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 168 ANLYGLNLNY 177  
 DB |||||  
 DB 28 ANLYGLNLNY 37

RESULT 7  
 AAR40100  
 ID AAR40100 standard; peptide; 53 AA.  
 XX  
 AC AAR40100;  
 XX  
 DT 04-FEB-1994 (first entry)  
 XX  
 DE Hib OMP P1-P2 hybrid peptide 2P6-1P13.  
 XX  
 KW Haemophilus influenzae; type b; Hib; outer membrane protein; P1; P2;  
 KW P6; vaccine; antibody; detection; lipoglycopeptide conjugate;  
 KW immunogen.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..15  
 FT /label= C-P1  
 FT Region 16..53  
 FT /label= CHIBP2  
 XX  
 PN WO9315205-A.  
 XX  
 PD 05-AUG-1993.  
 XX  
 PF 03-FEB-1993; 93WO-CA00041.  
 XX  
 PR 03-FEB-1992; 92GB-0002219.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 XX  
 PI Chong P, Kandil A, Klein MH, Sia C;  
 XX  
 DR WPI; 1993-258681/32.  
 XX  
 PT Synthetic Haemophilus influenzae conjugate vaccine - comprising  
 PT T-helper cell determinants and B-cell epitope(s) linked to  
 PT synthetic oligo:saccharide(s)  
 XX  
 PS Table 11; Page 59; 99pp; English.  
 XX  
 CC The sequences given in AAR40053-101 are peptide fragments derived from  
 CC the Haemophilus influenzae type b (Hib) outer membrane proteins P1,

XX  
 DR WPI; 1993-258681/32.  
 XX  
 PT Synthetic Haemophilus influenzae conjugate vaccine - comprising  
 PT T-helper cell determinants and B-cell epitope(s) linked to  
 PT synthetic oligo:saccharide(s)  
 XX  
 PS Table 1; Page 48; 99pp; English.  
 XX  
 CC The sequences given in AAR40053-101 are peptide fragments derived from  
 CC the Haemophilus influenzae type b (Hib) outer membrane proteins P1,  
 CC P2 and P6. These peptides may be used in a vaccine against Hib  
 CC infection and antibodies against these peptides may be used in test  
 CC kits to detect H. influenzae in a sample. The vaccine may further  
 CC comprise an immunogenic or immunostimulatory molecule or the peptides  
 CC may be modified with lipids, or linked to synthetic PRP as synthetic  
 CC lipoglycopeptide conjugates to produce alternative vaccines.  
 XX  
 XX Sequence 39 AA;  
 SQ

Query Match 5.6%; Score 10; DB 14; Length 39;  
 Best Local Similarity 100.0%; Pred. No. 0.018;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 168 ANLYGLNLNY 177  
 DB |||||  
 DB 28 ANLYGLNLNY 37

RESULT 7  
 AAR40100  
 ID AAR40100 standard; peptide; 53 AA.  
 XX  
 AC AAR40100;  
 XX  
 DT 04-FEB-1994 (first entry)  
 XX  
 DE Hib OMP P1-P2 hybrid peptide 2P6-1P13.  
 XX  
 KW Haemophilus influenzae; type b; Hib; outer membrane protein; P1; P2;  
 KW P6; vaccine; antibody; detection; lipoglycopeptide conjugate;  
 KW immunogen.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..15  
 FT /label= C-P1  
 FT Region 16..53  
 FT /label= CHIBP2  
 XX  
 PN WO9315205-A.  
 XX  
 PD 05-AUG-1993.  
 XX  
 PF 03-FEB-1993; 93WO-CA00041.  
 XX  
 PR 03-FEB-1992; 92GB-0002219.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 XX  
 PI Chong P, Kandil A, Klein MH, Sia C;  
 XX  
 DR WPI; 1993-258681/32.  
 XX  
 PT Synthetic Haemophilus influenzae conjugate vaccine - comprising  
 PT T-helper cell determinants and B-cell epitope(s) linked to  
 PT synthetic oligo:saccharide(s)  
 XX  
 PS Table 11; Page 59; 99pp; English.  
 XX  
 CC The sequences given in AAR40053-101 are peptide fragments derived from  
 CC the Haemophilus influenzae type b (Hib) outer membrane proteins P1,

CC P2 and P6. These peptides may be used in a vaccine against Hib  
 CC infection and antibodies against these peptides may be used in test  
 CC kits to detect H. influenzae in a sample. The vaccine may further  
 CC comprise an immunogenic or immunostimulatory molecule or the peptides  
 CC may be modified with lipids, or linked to synthetic PRP as synthetic  
 CC lipoglycopeptide conjugates to produce alternative vaccines.  
 XX  
 XX Sequence 53 AA;

Query Match 5.6%; Score 10; DB 14; Length 53;  
 Best Local Similarity 100.0%; Pred. No. 0.024; 0; Indels 0; Gaps 0;  
 Matches 10; Conservative 0; Mismatches 0;

QY 168 ANLYGLNLNY 177  
 DB 42 ANLYGLNLNY 51  
 |||||

RESULT 8  
 AAR40101  
 ID AAR40101 standard; peptide; 54 AA.  
 XX  
 AC AAR40101;  
 XX  
 DT 04-FEB-1994 (first entry)  
 XX  
 DE Hib OMP P1-P2 hybrid peptide CP2-1P13.  
 XX  
 KW Haemophilus influenzae; type b; Hib; outer membrane protein; P1; P2;  
 KW P6; vaccine; antibody; detection; lipoglycopeptide conjugate;  
 KW immunogen.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..16  
 FT /label= C-P1  
 FT Region 17..53  
 FT /label= CHIBP2  
 FT  
 XX WO9315205-A.  
 XX  
 XX 05-AUG-1993.  
 XX  
 XX 03-FEB-1993; 93WO-CA00041.  
 XX  
 XX 03-FEB-1992; 92GB-0002219.  
 XX  
 XX (CONN-) CONNAUGHT LAB LTD.  
 XX  
 XX Chong P, Kandil A, Klein MH, Sia C;  
 XX WPI; 1993-258681/32.  
 XX  
 XX Synthetic Haemophilus influenzae conjugate vaccine - comprising  
 XX T-helper cell determinants and B-cell epitope(s) linked to  
 XX synthetic oligo:saccharide(s)  
 XX  
 XX Table 11; Page 59; 99pp; English.  
 XX  
 XX The sequences given in AAR40053-101 are peptide fragments derived from  
 XX the Haemophilus influenzae type b (Hib) outer membrane proteins P1,  
 XX P2 and P6. These peptides may be used in a vaccine against Hib  
 XX infection and antibodies against these peptides may be used in test  
 XX kits to detect H. influenzae in a sample. The vaccine may further  
 XX comprise an immunogenic or immunostimulatory molecule or the peptides  
 XX may be modified with lipids, or linked to synthetic PRP as synthetic  
 XX lipoglycopeptide conjugates to produce alternative vaccines.  
 XX  
 XX Sequence 54 AA;

Query Match 5.6%; Score 10; DB 14; Length 54;  
 Best Local Similarity 100.0%; Pred. No. 0.024;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 168 ANLYGLNLNY 177  
 DB 43 ANLYGLNLNY 52  
 |||||

RESULT 9  
 AAR40099  
 ID AAR40099 standard; peptide; 55 AA.  
 XX  
 AC AAR40099;  
 XX  
 DT 04-FEB-1994 (first entry)  
 XX  
 DE Hib OMP P1-P2 hybrid peptide 2P2-1P13.  
 XX  
 KW Haemophilus influenzae; type b; Hib; outer membrane protein; P1; P2;  
 KW P6; vaccine; antibody; detection; lipoglycopeptide conjugate;  
 KW immunogen.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..17  
 FT /label= C-P1  
 FT Region 18..55  
 FT /label= CHIBP2  
 FT  
 XX WO9315205-A.  
 XX  
 XX 05-AUG-1993.  
 XX  
 XX 03-FEB-1993; 93WO-CA00041.  
 XX  
 XX 03-FEB-1992; 92GB-0002219.  
 XX  
 XX (CONN-) CONNAUGHT LAB LTD.  
 XX  
 XX Chong P, Kandil A, Klein MH, Sia C;  
 XX WPI; 1993-258681/32.  
 XX  
 XX Synthetic Haemophilus influenzae conjugate vaccine - comprising  
 XX T-helper cell determinants and B-cell epitope(s) linked to  
 XX synthetic oligo:saccharide(s)  
 XX  
 XX Table 11; Page 59; 99pp; English.  
 XX  
 XX The sequences given in AAR40053-101 are peptide fragments derived from  
 XX the Haemophilus influenzae type b (Hib) outer membrane proteins P1,  
 XX P2 and P6. These peptides may be used in a vaccine against Hib  
 XX infection and antibodies against these peptides may be used in test  
 XX kits to detect H. influenzae in a sample. The vaccine may further  
 XX comprise an immunogenic or immunostimulatory molecule or the peptides  
 XX may be modified with lipids, or linked to synthetic PRP as synthetic  
 XX lipoglycopeptide conjugates to produce alternative vaccines.  
 XX  
 XX Sequence 55 AA;

Query Match 5.6%; Score 10; DB 14; Length 55;  
 Best Local Similarity 100.0%; Pred. No. 0.025;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 ANLYGLNLNY 177  
 DB 44 ANLYGLNLNY 53  
 |||||

RESULT 10  
 AAB96665  
 ID AAB96665 standard; Protein; 636 AA.  
 XX

```

AC AAB96665;
XX
DT 29-OCT-2001 (first entry)
XX
DE Putative P. abyssi ATPase subunit of ABC transporter #18.
XX
KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS Pyrococcus abyssi.
XX
PN FR2792651-A1.
XX
PD 27-OCT-2000.
XX
PF 21-APR-1999; 99FR-0005034.
XX
PR 21-APR-1999; 99FR-0005034.
XX
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI (IPRE-) IPREMER INST FR RECH EXPL MER.
XX
PI Forterre P, Thierry JC, Priour D, Dietrich J, Lecompte O;
PI Querellou J, Weissenbach J, Saurin W, Heilig R;
XX
DR WPI; 2001-126236/14.
XX
XX New nucleotide sequences isolated from Pyrococcus abyssi encode
PT proteins useful in industry -
XX
PS Claim 7; Pages 1419-1421; 1657pp; French.
XX
CC The present invention relates to the genomic sequence of Pyrococcus
CC abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
CC a hyperthermophilic archaeon, which is isolated from deep-sea
CC hydrothermal vents. The present sequence is one such P. abyssi protein.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to
CC 110 degrees centigrade.
CC Note: This patent is in the same patent family as WO2000065062, which
CC contains additional sequences as shown in AAB99132-AAB99143,
CC AAH75903-AAH75920 and AAG66436.
XX
SQ Sequence 636 AA;
Query Match 4.5%; Score 8; DB 22; Length 636;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 142 KKKHFVET 149
Db 322 KKKHFVET 329
|||||

RESULT 11
AAY33535
ID AAY33535 standard; Protein: 61 AA.
XX
AC AAY33535;
XX
DT 20-JAN-2000 (first entry)
XX
DE Tobacco DNA binding protein EREBP2 AP2 domain.
XX
KW Environmental stress tolerance; plant; binding protein; DNA regulation;
KW CBF-1; C-repeat/DRE binding factor; CCG regulatory sequence; COR; cold;
KW cold-related gene; drought; high salinity; tissue-specific promoter;
KW AP2 domain; tobacco; EREBP2.
XX
OS Nicotiana tabacum.
XX
PN WO9938977-A2.
XX
PD 05-AUG-1999.

```

```

XX 28-JAN-1999; 99WO-US01895.
XX
XX 03-FEB-1998; 98US-0017575.
XX 03-FEB-1998; 98US-0017816.
XX 03-FEB-1998; 98US-0018227.
XX 03-FEB-1998; 98US-0018233.
XX 03-FEB-1998; 98US-0018234.
XX 03-FEB-1998; 98US-0018235.
XX 23-NOV-1998; 98US-0198119.
XX
XX (MEND-) MENDEL BIOTECHNOLOGY INC.
XX (UNMS ) UNIV MICHIGAN STATE.
XX
XX Stockinger EJ, Jaglo-Ottosen K, Zarka D, Gilmour SJ, Jiang C;
XX Fromm M, Thomashow MF;
XX
XX WPI; 1999-561312/47.
XX
XX Environmental stress tolerance gene binding proteins useful for
PT altering plant stress tolerance -
XX
XX Example 1D; Fig 2D; 252pp; English.
XX
XX This invention describes novel binding proteins other than CBF-1
XX (C-repeat/DRE binding factor) in isolated form which comprise a
XX consensus sequence capable of binding to a CCG regulatory sequence.
XX The binding proteins are capable of binding to a DNA regulatory
XX sequence, which regulates expression of one or more environmental
XX stress tolerance genes, especially COR (cold-related) genes.
XX Environmental stress may be, e.g. cold temperatures, drought and high
XX salinity. Plants transformed with the binding protein (or sequences
XX encoding it) can have altered environmental stress tolerance. The
XX binding protein coding sequences can be under the control of
XX tissue-specific promoters. This sequence represents the tobacco DNA
XX binding protein EREBP2 AP2 domain which is used to describe the method
XX of the invention.
XX
XX Sequence 61 AA;
Query Match 3.9%; Score 7; DB 20; Length 61;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 95 LAYDKAA 101
Db 40 LAYDKAA 46
|||||

RESULT 12
AAY31478
ID AAY31478 standard; peptide: 68 AA.
XX
XX AAY31478;
XX
XX 29-OCT-1999 (first entry)
XX
XX A. thaliana RAP2.4 EREBP-like subclass AP2 domain.
XX
XX ADC gene; AP2 domain containing gene; regulatory gene; APETALA2;
XX seed mass modulation; genetic engineering; transgenic plant.
XX
XX Arabidopsis thaliana.
XX
XX WO9941974-A1.
XX
XX 26-AUG-1999.
XX
XX 17-FEB-1999; 99WO-US03429.
XX
XX 19-FEB-1998; 98US-0026039.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX

```

XX Jofuku KD, Okamuro JK;  
 XX WPI; 1999-518486/43.  
 XX Novel methods for modulating seed mass and other plant traits using  
 XX new expression cassettes containing a plant promoter  
 XX Example 4; Page 84; 104pp; English.  
 XX The invention provides novel methods for controlling seed size and total  
 CC seed protein using ADC (AP2 domain containing) gene, which is a plant  
 CC regulatory gene, over expression and antisense gene constructs. The  
 CC method of modulating seed mass in a plant comprises: providing a first  
 CC plant comprising a recombinant expression cassette containing an ADC  
 CC nucleic acid linked to a plant promoter; selfing the first plant or  
 CC crossing the first plant with a second plant, thereby producing a  
 CC plurality of seeds; and selecting seeds with altered mass. The methods  
 CC of the invention can be used to enhance or increase endogenous gene  
 CC expression. Enhanced ADC expression leads to smaller seeds or seedless  
 CC fruit. The methods can be used to produce a broad range of transgenic  
 CC plants. Increasing seed size, amino acid content, and oil content is  
 CC desirable in crop plants for human or animal consumption, e.g. soybean,  
 CC rice, wheat, corn, rye, etc.. Decreasing seed size is useful in plants  
 CC grown for their fruit and where large seeds are undesirable, e.g.  
 CC cucumbers, tomatoes, cherries, melons. Sequences AAY31456-496 represent  
 CC AP2 domain fragments of various polypeptides.  
 XX Sequence 68 AA;

Query Match 3.9%; Score 7; DB 20; Length 68;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 LAYDKAA 101  
 Db 41 LAYDKAA 47

RESULT 13  
 AAB25837  
 ID AAB25837 standard; Protein; 68 AA.  
 XX AC AAB25837;  
 XX 15-DEC-2000 (first entry)  
 XX DE AP2 domain amino acid sequence of Arabidopsis RAP2 protein.  
 XX KW Apetela; AP2 domain; canola; soybean; ANT; RAP; EREBP; flower; seedless;  
 XX KW fruit; cherry; melon; tomato; AP2 domain containing; ADC;  
 XX KW transgenic plant.  
 XX OS Arabidopsis sp.  
 XX PN US6093874-A.  
 XX PD 25-JUL-2000.  
 XX PF 15-AUG-1997; 97US-0912272.  
 XX PR 20-AUG-1996; 96US-0700152.  
 XX PR 20-JUN-1997; 97US-0879827.  
 XX PA (REGC ) UNIV CALIFORNIA.  
 XX PI Okamuro JK, Jofuku KD;  
 XX WPI; 2000-514122/46.  
 XX Modulating seed traits useful for creating transgenic plants with  
 PT altered seed size or protein content, by providing a plant with a  
 PT recombinant expression cassette containing an ADC (AP2 domain

PT containing) nucleic acid linked to a promoter -  
 XX Disclosure; Figure 4; 60pp; English.  
 XX This invention relates to a method for modulating seed traits in a  
 CC soybean or canola plant comprising providing a plant having a  
 CC recombinant expression cassette containing an ADC (AP2 domain containing)  
 CC nucleic acid linked to a plant promoter. ADC gene sequences are  
 CC represented by AAA96979-A96981. The AP2 gene (APETALA) was originally  
 CC isolated from Arabidopsis sp. and AP2 domain containing genes include  
 CC RAP2 (related to AP2) ANT and EREBPs. The AP2 gene is a floral homeotic  
 CC gene. Each of these proteins contain AP2 like domains. Sequences  
 CC AAB25813-B25858 represent protein fragments and peptides from ADC  
 CC proteins such as RAP2.7. Sequences AAA96982-A97035 represent PCR primers  
 CC used to isolate the ADC genes of the invention. The method is used for  
 CC modulating seed traits in canola and soybean plants and for creating  
 CC transgenic plants. The nucleic acids can be used to confer desired traits  
 CC on essentially any plant by increasing or decreasing gene expression.  
 CC These may be used to increase seed size, protein, amino acid or oil  
 CC content in crop plants in which seeds are used directly for animal or  
 CC human consumption or for industrial purposes, such as soybean, canola and  
 CC other grains. These may be also be used to decrease seed size or  
 CC producing seedless varieties in plants grown for their fruit and in which  
 CC large seeds may be undesirable, such as cherries, melons or tomatoes.  
 CC Primers, which specifically amplify AP2 domains of the genes, are  
 CC particularly useful for identification of particular ADC  
 CC polynucleotides.  
 XX Sequence 68 AA;

Query Match 3.9%; Score 7; DB 21; Length 68;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 LAYDKAA 101  
 Db 41 LAYDKAA 47

RESULT 14  
 AAY31468  
 ID AAY31468 standard; peptide; 69 AA.  
 XX AC AAY31468;  
 XX DT 29-OCT-1999 (first entry)  
 XX DE N. tabacum EREBP-2 EREBP-like subclass AP2 domain.  
 XX KW ADC gene; AP2 domain containing gene; regulatory gene; APETALA2;  
 XX KW seed mass modulation; genetic engineering; transgenic plant.  
 XX OS Nicotiana tabacum.  
 XX PN WO9941974-A1.  
 XX PD 26-AUG-1999.  
 XX PF 17-FEB-1999; 99WO-US03429.  
 XX PR 19-FEB-1998; 98US-0026039.  
 XX PA (REGC ) UNIV CALIFORNIA.  
 XX PI Jofuku KD, Okamuro JK;  
 XX WPI; 1999-518486/43.  
 XX Novel methods for modulating seed mass and other plant traits using  
 PT new expression cassettes containing a plant promoter  
 XX Example 4; Page 79-80; 104pp; English.



PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 21-MAY-1999; 99US-0135629.  
PR 24-MAY-1999; 99US-0136021.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139453.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 16-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.

PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 3.9%; Score 7; DB 21; Length 79;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 NIKGLLL 157  
| | | | |  
Db 5 NIKGLLL 11

RESULT 17  
AAG92876  
ID AAG92876 standard; Protein; 93 AA.  
XX AC AAG92876;  
XX DT 26-SEP-2001 (first entry)  
XX DE C glutamicum protein fragment SEQ ID NO: 6630.

XX Corynebacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis.  
XX OS Corynebacterium glutamicum.

PN EP1108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

XX 07-APR-2000; 2000JP-0159162.

XX 03-AUG-2000; 2000JP-0280988.

XX (KYOW ) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.

XX N-PSDB; AAH68095.

PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
PT  
XX  
XX

PS Claim 17; SEQ ID NO: 6630; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein  
CC sequences from the Corynebacterium glutamicum glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of corynebacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Corynebacterium glutamicum, and identifying a homologue of a gene derived  
CC from corynebacterium. Corynebacterium bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention.  
CC

CC Note: The exemplification data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.

XX Sequence 93 AA;

Query Match 3.9%; Score 7; DB 22; Length 93;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 IKGLLLV 158  
| | | | |  
Db 82 IKGLLLV 88

RESULT 18  
AAG32798  
ID AAG32798 standard; Protein; 95 AA.  
XX AC AAG32798;  
XX DT 17-OCT-2000 (first entry)

XX Zea mays protein fragment SEQ ID NO: 39637.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence; corn.

XX Zea mays subsp. mays.

XX PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 04-MAY-1999; 99US-0132484.

XX 05-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 07-MAY-1999; 99US-0132487.

XX 11-MAY-1999; 99US-0132863.

XX 14-MAY-1999; 99US-0134256.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 18-MAY-1999; 99US-0134370.

XX 19-MAY-1999; 99US-0134768.

XX 19-MAY-1999; 99US-0134941.

XX 20-MAY-1999; 99US-0135124.

XX 21-MAY-1999; 99US-0135353.

XX 24-MAY-1999; 99US-0135629.

XX 25-MAY-1999; 99US-0136021.

XX 27-MAY-1999; 99US-0136392.

XX 28-MAY-1999; 99US-0136782.

XX 01-JUN-1999; 99US-0137222.

XX 03-JUN-1999; 99US-0137528.



PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 17-JUN-1999; 99US-0139494.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match  
Best Local Similarity

3.9%  
100.0%

Score 7; DB 21;  
Pred. No. 44;  
Length 95;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 NIKGLLL 157  
|||||  
Db 21 NIKGLLL 27

RESULT 19  
AAU54058  
ID AAU54058 standard; Protein: 109 AA.  
XX AC  
XX AAU54058;  
XX DT 27-FEB-2002 (first entry)  
XX DE Propionibacterium acnes immunogenic protein #14954.  
XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.  
XX OS  
XX Propionibacterium acnes.  
XX PN W0200181581-A2.  
XX PD 01-NOV-2001.  
XX PF 20-APR-2001; 2001WO-US12865.  
XX PR 21-APR-2000; 2000US-199047P.  
XX PR 02-JUN-2000; 2000US-208841P.  
XX PR 07-JUL-2000; 2000US-216747P.  
XX PA (CORI-) CORIXA CORP.  
XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX N-PSDB; AAS59563.  
XX WPI: 2001-616774/71.  
XX DR N-PSDB; AAS59563.  
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris.  
XX Example 1; SEQ ID No 15253; 1069pp; English.  
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 109 AA;

Query Match 3.9%; Score 7; DB 22; Length 109;  
Best Local Similarity 100.0%; Pred. No. 51;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 106 LSASIPD 112  
|||||  
Db 68 LSASIPD 74

RESULT 20  
AAU48379  
ID AAU48379 standard; Protein: 113 AA.  
XX AC  
XX AAU48379;  
XX DT 27-FEB-2002 (first entry)  
XX DE Propionibacterium acnes immunogenic protein #9275.  
XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.  
XX OS  
XX Propionibacterium acnes.  
XX PN W0200181581-A2.  
XX PD 01-NOV-2001.  
XX PF 20-APR-2001; 2001WO-US12865.  
XX PR 21-APR-2000; 2000US-199047P.  
XX PR 02-JUN-2000; 2000US-208841P.  
XX PR 07-JUL-2000; 2000US-216747P.  
XX PA (CORI-) CORIXA CORP.  
XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX N-PSDB; AAS59542.  
XX WPI: 2001-616774/71.  
XX DR N-PSDB; AAS59542.  
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris.  
XX Example 1; SEQ ID No 9574; 1069pp; English.  
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 113 AA;

Query Match 3.9%; Score 7; DB 22; Length 113;  
Best Local Similarity 100.0%; Pred. No. 52;





XX AAM60957;  
 AC 13-OCT-1998 (first entry)  
 DT Streptococcus pneumoniae encoded polypeptide.  
 XX coding region; ORF; open reading frame; antibacterial;  
 KW infection; prevention; meningitis.  
 KW Streptococcus pneumoniae.  
 OS WO9819689-A1.  
 XX 14-MAY-1998.  
 PD 27-OCT-1997; 97WO-US19226.  
 PF 01-NOV-1996; 96US-0029930.  
 XX (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;  
 PI Reid RH, Zarfos FN;  
 PI WPI; 1998-286586/25.  
 DR N-PSDB; AAV37356.  
 XX New isolated nucleic acids from Streptococcus pneumoniae - useful,  
 PT e.g. for identifying anti-bacterial(s) for treatment and prevention  
 PT of meningitis  
 XX Claim 11; Page 57; 130pp; English.  
 XX The sequence is that of the polypeptide encoded by a region isolated  
 CC from S. pneumoniae which shows homology to ISL2 protein.  
 CC The protein, or agonists of it, may be useful as an antibacterial for  
 CC treatment or prevention of infection, specifically caused by S.  
 CC pneumoniae (particularly meningitis) but possibly also Helicobacter  
 CC pylori (ulcers and gastric cancer). It may be of particular  
 CC use before insertion of an in-dwelling device or any other  
 CC invasive procedure. The protein, or nucleic acid encoding  
 CC it, can also be used in vaccines to induce a cellular  
 CC and/or humoral immune response, or to screen for other  
 CC antibacterials. The DNA may also contain flanking sequences  
 CC that are potential sources of control elements for bacterial  
 CC gene expression. Detecting a sequence encoding the protein  
 CC can be used diagnostically, e.g. to detect a mutation for  
 CC serotyping or classifying infectious agents.  
 XX  
 XX Sequence 206 AA;  
 Query Match 3.9%; Score 7; DB 19; Length 206;  
 Best Local Similarity 100.0%; Pred. No. 92;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 102 SKTHLSA 108  
 Db 40 SKTHLSA 46  
 RESULT 24  
 ABP26302  
 ID: ABP26302 standard; Protein; 207 AA.  
 XX  
 XX ABP26302;  
 XX  
 XX 02-JUL-2002 (first entry)  
 XX Streptococcus polypeptide SEQ ID NO 1780.  
 DE Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW

KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX Streptococcus agalactiae.  
 XX WO200234771-A2.  
 PN 02-MAY-2002.  
 XX 29-OCT-2001; 2001WO-GB04789.  
 PF 27-OCT-2000; 2000GB-0026333.  
 PR 24-NOV-2000; 2000GB-0028727.  
 PR 07-MAR-2001; 2001GB-0005640.  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;  
 PI Tettelin H;  
 DR WPI; 2002-352536/38.  
 DR N-PSDB; ABN66933.  
 XX New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -  
 XX Claim 1; Page 3328; 4525pp; English.  
 XX The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.  
 XX Sequence 207 AA;  
 Query Match 3.9%; Score 7; DB 23; Length 207;  
 Best Local Similarity 100.0%; Pred. No. 93;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 102 SKTHLSA 108  
 Db 181 SKTHLSA 187  
 RESULT 25  
 AAE02538  
 ID AAE02538 standard; Protein; 218 AA.  
 XX  
 XX AAE02538;  
 AC  
 XX 10-AUG-2001 (first entry)  
 DT  
 XX A. thaliana transcription factor G26.  
 DE  
 XX Plant transcription factor; phenotype; sugar sensing characteristic;  
 KW transgenic plant; plant yield; growth; germination; photosynthesis;  
 KW glyoxylate metabolism; respiration; pathogen response; wounding response;  
 KW cell cycle regulation; pigmentation; flowering; senescence; physiology;

```

KW storage organ; metabolism.
XX Arabidopsis thaliana.
XX
XX Key Location/Qualifiers
XX Domain 67..134
XX /note= "Conserved domain"
XX
XX WO200135725-A1.
XX
XX 25-MAY-2001.
XX
XX 14-NOV-2000; 2000WO-US31414.
XX
XX 17-NOV-1999; 99US-0166228.
XX
XX 17-APR-2000; 2000US-0197899.
XX
XX 22-AUG-2000; 2000US-0227439.
XX
XX (MEND-) MENDEL BIOTECHNOLOGY INC.
XX (JIAN/) JIANG C.
XX (HEAR/) HEARD J.
XX (PINE/) PINEDA O.
XX (PILG/) PILGRIM M.
XX (ADAM/) ADAM L.
XX (RIEC/) RIECHMANN J L.
XX (YUGG/) YU G.
XX (SAMA/) SAMARA R.
XX
XX JIANG C, Heard J, Pineda O, Pilgrim M, Adam L, Riechmann JL;
XX Yu G, Samaha R;
XX
XX WPI; 2001-335977/35.
XX N-PSDB; AAD06639.
XX
XX Nucleic acids encoding plant transcription factor polypeptides, useful
XX for altering the sugar sensing characteristics of plants and increasing
XX yield, e.g. corn, potato and cotton plants -
XX
XX Claim 4; Page 57-58; 151pp; English.
XX
XX The patent relates to polynucleotides encoding 35 plant transcription
XX factors which may be used to modify phenotype associated with a plant's
XX sugar sensing characteristics and increasing yield when their expression
XX level is altered. Sugars are central regulatory molecules that control
XX aspects of physiology, metabolism and development. Therefore the cDNAs
XX and proteins of the invention are useful for modifying the growth and
XX germination rates of plants, photosynthesis, glyoxylate metabolism,
XX respiration, starch and sucrose synthesis and degradation, pathogen
XX response, wounding response, cell cycle regulation, pigmentation,
XX flowering and senescence of plants and for modifying sink-source
XX relationships in seeds, tubers, roots, and other storage organs leading
XX to an increase in yield. The transcription factor polynucleotides and
XX polypeptides may be used to alter the structure and developmental
XX characteristics of plants such as soybean, wheat, corn, potato, cotton,
XX rice, oilseed, sunflower, alfalfa, sugarcane, turf, banana, blackberry,
XX blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower,
XX coffee, cucumber, eggplant, grapes, honey dew, lettuce, mango, melon,
XX onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn,
XX tobacco, tomato, watermelon, rosaceous fruits and/or vegetable brassicas.
XX The present sequence is an Arabidopsis thaliana transcription factor.
XX
XX Sequence 218 AA;
XX
XX Query Match 3.9%; Score 7; DB 22; Length 218;
XX Best Local Similarity 100.0%; Pred. No. 97;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 95 LAYDKAA 101
XX |
XX Db 107 LAYDKAA 113
XX
XX RESULT 26
XX
XX storage organ; metabolism; protein; 219 AA.
XX
XX AAB32762;
XX
XX 25-JAN-2001 (first entry)
XX
XX Eucalyptus grandis transcription factor protein sequence #220.
XX
XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
XX poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
XX basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
XX homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
XX type 2 Cys2His2; CCAAT box element; MYB.
XX
XX Eucalyptus grandis.
XX
XX WO200053724-A2.
XX
XX 14-SEP-2000.
XX
XX 09-MAR-2000; 2000WO-US06112.
XX
XX 11-MAR-1999; 99US-0266513.
XX
XX 18-AUG-1999; 99US-0149485.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Wood M, McGrath A, Shenk MA, Glenn M;
XX
XX WPI; 2000-579369/54.
XX
XX New isolated polynucleotide encoding a plant transcription factor for
XX producing a plant e.g. a woody plant, preferably eucalyptus or pine,
XX having modified gene expression or modified activity of a polypeptide
XX
XX Claim 8; Page 293; 747pp; English.
XX
XX The present invention relates to novel plant transcription factors from
XX Eucalyptus grandis or Pinus radiata. The present sequence is one such
XX transcription factor. The transcription factor may be used to produce a
XX plant having modified gene expression such as a woody plant e.g. a
XX eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or
XX to modify the activity of a polypeptide in a plant. The transcription
XX factors of the present invention are members from the following families
XX of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic
XX helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain
XX zipper, LIM domain, AP2 and EREBs, zinc finger domains of type 2
XX Cys2His2, CCAAT box elements and MYB.
XX
XX Sequence 219 AA;
XX
XX Query Match 3.9%; Score 7; DB 21; Length 219;
XX Best Local Similarity 100.0%; Pred. No. 98;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 95 LAYDKAA 101
XX |
XX Db 65 LAYDKAA 71
XX
XX RESULT 27
XX AAY21977
XX ID AAY21977 standard; Protein; 230 AA.
XX
XX AAY21977;
XX
XX 06-SEP-1999 (first entry)
XX
XX Sark protein fragment.
XX

```

KW Senescence-associated receptor-like protein kinase; sark; sam gene; sag;  
 KW senescence-associated gene; plant senescence; promoter; pharmaceutical;  
 KW plant maturation; S-adenosyl methionine; flower; fruit development.

OS Phaseolus vulgaris.

XX WO929159-A1.

XX 17-JUN-1999.

PD 08-DEC-1998; 98WO-US25799.

PF 08-DEC-1997; 97US-0067898.

XX (VITA-) VITALITY BIOTECHNOLOGIES INC.

XX Cepstein S, Hajjuje T, Rosner A;

XX WPI; 1999-404873/34.

DR N-PSDB; AAX81126.

XX DNA encoding senescence-associated genes for a senescence

PT receptor-like protein kinase

XX Disclosure; Fig 2; 70pp; English.

CC The invention relates to a senescence-associated receptor-like protein  
 CC kinase (sark) gene. The sark gene is a senescence-associated gene (sag)  
 CC and is expressed early in the plant senescence process. The sark gene  
 CC promoter is useful for driving expression of foreign genes having a  
 CC desired product, such as a pharmaceutical, during the process of plant  
 CC maturation. The sark gene promoter can be used to drive expression of  
 CC resistance genes against pathogens or pests during senescence when the  
 CC plant is particularly susceptible to infection or infestation. The sark  
 CC gene promoter may also be used to drive expression of a gene encoding an  
 CC inhibitor of senescence. Plant senescence may be inhibited by use of  
 CC antisense sark constructs. Over expression of the sag genes, using the  
 CC sark or sam (S-adenosyl methionine) gene promoters is useful for  
 CC induction of early senescence. This is useful to obtain flower or fruit  
 CC development prior to specific pest onset, prior to undesirable cross-  
 CC fertilization from related crops, at a specific time during storage or  
 CC retail, or to avoid development of plant structures that are not of  
 CC agronomic importance. The present sequence represents a region of the  
 CC sark protein selected for expression in E. coli.

SQ Sequence 230 AA;

Query Match 3.9%; Score 7; DB 20; Length 230;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 LSASIPD 112

DB 144 LSASIPD 150

RESULT 28

AAG38454

ID AAG38454 standard; Protein; 313 AA.

XX AAG38454;

AC AAG38454;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 47443.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000. 2000EP-0301439.  
 PD 25-FEB-1999; 99US-0121825.  
 XX 05-MAR-1999; 99US-0123180.  
 XX 09-MAR-1999; 99US-0123548.  
 XX 23-MAR-1999; 99US-0125788.  
 XX 25-MAR-1999; 99US-0126264.  
 XX 29-MAR-1999; 99US-0126785.  
 XX 01-APR-1999; 99US-0127462.  
 XX 06-APR-1999; 99US-0128234.  
 XX 08-APR-1999; 99US-0128714.  
 XX 16-APR-1999; 99US-0129845.  
 XX 19-APR-1999; 99US-0130077.  
 XX 21-APR-1999; 99US-0130449.  
 XX 23-APR-1999; 99US-0130510.  
 XX 28-APR-1999; 99US-0130891.  
 XX 30-APR-1999; 99US-0131449.  
 XX 04-MAY-1999; 99US-0132048.  
 XX 05-MAY-1999; 99US-0132484.  
 XX 06-MAY-1999; 99US-0132485.  
 XX 07-MAY-1999; 99US-0132487.  
 XX 11-MAY-1999; 99US-0132863.  
 XX 14-MAY-1999; 99US-0134256.  
 XX 14-MAY-1999; 99US-0134218.  
 XX 14-MAY-1999; 99US-0134219.  
 XX 14-MAY-1999; 99US-0134221.  
 XX 18-MAY-1999; 99US-0134370.  
 XX 19-MAY-1999; 99US-0134768.  
 XX 20-MAY-1999; 99US-0134941.  
 XX 21-MAY-1999; 99US-0135124.  
 XX 24-MAY-1999; 99US-0135353.  
 XX 25-MAY-1999; 99US-0135629.  
 XX 27-MAY-1999; 99US-0136021.  
 XX 28-MAY-1999; 99US-0136392.  
 XX 01-JUN-1999; 99US-0136782.  
 XX 03-JUN-1999; 99US-0137222.  
 XX 04-JUN-1999; 99US-0137528.  
 XX 07-JUN-1999; 99US-0137724.  
 XX 08-JUN-1999; 99US-0138094.  
 XX 10-JUN-1999; 99US-0138540.  
 XX 14-JUN-1999; 99US-0138847.  
 XX 16-JUN-1999; 99US-0139119.  
 XX 17-JUN-1999; 99US-0139452.  
 XX 18-JUN-1999; 99US-0139453.  
 XX 18-JUN-1999; 99US-0139492.  
 XX 18-JUN-1999; 99US-0139454.  
 XX 18-JUN-1999; 99US-0139455.  
 XX 18-JUN-1999; 99US-0139456.  
 XX 18-JUN-1999; 99US-0139457.  
 XX 18-JUN-1999; 99US-0139458.  
 XX 18-JUN-1999; 99US-0139459.  
 XX 18-JUN-1999; 99US-0139460.  
 XX 18-JUN-1999; 99US-0139461.  
 XX 18-JUN-1999; 99US-0139462.  
 XX 18-JUN-1999; 99US-0139463.  
 XX 18-JUN-1999; 99US-0139750.  
 XX 21-JUN-1999; 99US-0139763.  
 XX 22-JUN-1999; 99US-0139817.  
 XX 23-JUN-1999; 99US-0139899.  
 XX 23-JUN-1999; 99US-0140353.  
 XX 24-JUN-1999; 99US-0140354.  
 XX 28-JUN-1999; 99US-0140695.  
 XX 29-JUN-1999; 99US-0140823.  
 XX 30-JUN-1999; 99US-0140991.  
 XX 01-JUL-1999; 99US-0141287.  
 XX 01-JUL-1999; 99US-0141842.  
 XX 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151920.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161932.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
PR 29-OCT-1999; 99US-0162145.

Query Match 3.9% Score 7; DB 21; Length 313;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 95 LAYDKAA 101  
| | | | |  
Db 168 LAYDKAA 174

## RESULT 29

AAG38787  
ID AAG38787 standard; Protein; 313 AA.  
XX AC AAG38787;  
XX DT 18-OCT-2000 (first entry)  
XX XX Arabidopsis thaliana protein fragment SEQ ID NO: 47902.  
DE DE Arabidopsis thaliana protein fragment SEQ ID NO: 47902.  
XX XX Arabidopsis thaliana protein fragment SEQ ID NO: 47902.  
KW KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EPI033405-A2.  
XX PD 06-SEP-2000.  
XX XX 25-FEB-2000; 2000EP-0301439.  
XX PF 25-FEB-1999; 99US-0121825.  
XX PR 05-MAR-1999; 99US-0123180.



PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.

```
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162143.
PR 29-OCT-1999; 99US-0162144.

Query Match 3.9%; Score 7; DB 21; Length 313;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 LAYDKAA 101
Db 168 LAYDKAA 174
|||||

RESULT 30
AAG38453
ID AAG38453 standard; Protein; 330 AA.
XX
AC AAG38453;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 47442.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
XX
PN EF1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
```

```
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149923.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.9%; Score 7; DB 21; Length 330;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 LAYDKAA 101
Db 185 LAYDKAA 191
|||||

RESULT 31
AAG38786
ID AAG38786 standard; Protein; 330 AA.
XX
AC AAG38786;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 47901.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
```

PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134321.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143342.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145087.  
PR 21-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 05-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.

```
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      3.9%; Score 7; DB 21; Length 330;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 LAYDKAA 101
Db 185 LAYDKAA 191

RESULT 32
AAG38452
ID AAG38452 standard; Protein; 334 AA.
XX
AC AAG38452;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 47441.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.

14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
14-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134768.
19-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137222.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139492.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144632.
20-JUL-1999; 99US-0144884.
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US-0145086.
21-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145085.
22-JUL-1999; 99US-0145087.
22-JUL-1999; 99US-0145089.
22-JUL-1999; 99US-0145192.
23-JUL-1999; 99US-0145145.
23-JUL-1999; 99US-0145218.
```

```
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145931.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146398.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147433.
PR 09-AUG-1999; 99US-0147933.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.

PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.9% Score 7; DB 21; Length 334;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 95 LAYDKAA 101
Db 189 LAYDKAA 195
|||||

RESULT 33
AAG38785
ID AAG38785 standard; Protein; 334 AA.
XX
AC AAG38785;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 47900.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
```

PR 21-MAY-1999; 990S-0135353.  
PR 24-MAY-1999; 990S-0135629.  
PR 25-MAY-1999; 990S-0136021.  
PR 27-MAY-1999; 990S-0136392.  
PR 28-MAY-1999; 990S-0136782.  
PR 01-JUN-1999; 990S-0137222.  
PR 03-JUN-1999; 990S-0137528.  
PR 04-JUN-1999; 990S-0137502.  
PR 07-JUN-1999; 990S-0137724.  
PR 08-JUN-1999; 990S-0138094.  
PR 10-JUN-1999; 990S-0138540.  
PR 10-JUN-1999; 990S-0138847.  
PR 14-JUN-1999; 990S-0139119.  
PR 16-JUN-1999; 990S-0139452.  
PR 16-JUN-1999; 990S-0139453.  
PR 17-JUN-1999; 990S-0139492.  
PR 18-JUN-1999; 990S-0139454.  
PR 18-JUN-1999; 990S-0139455.  
PR 18-JUN-1999; 990S-0139456.  
PR 18-JUN-1999; 990S-0139457.  
PR 18-JUN-1999; 990S-0139458.  
PR 18-JUN-1999; 990S-0139459.  
PR 18-JUN-1999; 990S-0139460.  
PR 18-JUN-1999; 990S-0139461.  
PR 18-JUN-1999; 990S-0139462.  
PR 18-JUN-1999; 990S-0139463.  
PR 18-JUN-1999; 990S-0139750.  
PR 18-JUN-1999; 990S-0139763.  
PR 21-JUN-1999; 990S-0139817.  
PR 22-JUN-1999; 990S-0139899.  
PR 23-JUN-1999; 990S-0140353.  
PR 23-JUN-1999; 990S-0140354.  
PR 24-JUN-1999; 990S-0140695.  
PR 28-JUN-1999; 990S-0140823.  
PR 29-JUN-1999; 990S-0140991.  
PR 30-JUN-1999; 990S-0141287.  
PR 01-JUL-1999; 990S-0141842.  
PR 01-JUL-1999; 990S-0142154.  
PR 02-JUL-1999; 990S-0142055.  
PR 06-JUL-1999; 990S-0142390.  
PR 08-JUL-1999; 990S-0142803.  
PR 09-JUL-1999; 990S-0142920.  
PR 12-JUL-1999; 990S-0142977.  
PR 13-JUL-1999; 990S-0143542.  
PR 14-JUL-1999; 990S-0143624.  
PR 15-JUL-1999; 990S-0144005.  
PR 16-JUL-1999; 990S-0144085.  
PR 16-JUL-1999; 990S-0144086.  
PR 19-JUL-1999; 990S-0144325.  
PR 19-JUL-1999; 990S-0144331.  
PR 19-JUL-1999; 990S-0144332.  
PR 19-JUL-1999; 990S-0144333.  
PR 19-JUL-1999; 990S-0144334.  
PR 19-JUL-1999; 990S-0144335.  
PR 20-JUL-1999; 990S-0144352.  
PR 20-JUL-1999; 990S-0144632.  
PR 20-JUL-1999; 990S-0144884.  
PR 21-JUL-1999; 990S-0144884.  
PR 21-JUL-1999; 990S-0145086.  
PR 21-JUL-1999; 990S-0145088.  
PR 22-JUL-1999; 990S-0145085.  
PR 22-JUL-1999; 990S-0145087.  
PR 22-JUL-1999; 990S-0145089.  
PR 22-JUL-1999; 990S-0145192.  
PR 23-JUL-1999; 990S-0145145.  
PR 23-JUL-1999; 990S-0145218.  
PR 23-JUL-1999; 990S-0145224.  
PR 26-JUL-1999; 990S-0145276.  
PR 27-JUL-1999; 990S-0145913.  
PR 27-JUL-1999; 990S-0145918.  
PR 27-JUL-1999; 990S-0145919.  
PR 28-JUL-1999; 990S-0145951.  
PR 02-AUG-1999; 990S-0146386.  
PR 02-AUG-1999; 990S-0146387.  
PR 02-AUG-1999; 990S-0146388.  
PR 02-AUG-1999; 990S-0146389.  
PR 03-AUG-1999; 990S-0147038.  
PR 04-AUG-1999; 990S-0147204.  
PR 04-AUG-1999; 990S-0147302.  
PR 05-AUG-1999; 990S-0147192.  
PR 05-AUG-1999; 990S-0147260.  
PR 06-AUG-1999; 990S-0147303.  
PR 06-AUG-1999; 990S-0147416.  
PR 09-AUG-1999; 990S-0147493.  
PR 09-AUG-1999; 990S-0147935.  
PR 10-AUG-1999; 990S-0148171.  
PR 11-AUG-1999; 990S-0148319.  
PR 12-AUG-1999; 990S-0148341.  
PR 13-AUG-1999; 990S-0148565.  
PR 13-AUG-1999; 990S-0148684.  
PR 16-AUG-1999; 990S-0149368.  
PR 17-AUG-1999; 990S-0149175.  
PR 18-AUG-1999; 990S-0149426.  
PR 20-AUG-1999; 990S-0149722.  
PR 20-AUG-1999; 990S-0149723.  
PR 20-AUG-1999; 990S-0149929.  
PR 23-AUG-1999; 990S-0149902.  
PR 23-AUG-1999; 990S-0149930.  
PR 25-AUG-1999; 990S-0150566.  
PR 26-AUG-1999; 990S-0150884.  
PR 27-AUG-1999; 990S-0151065.  
PR 27-AUG-1999; 990S-0151066.  
PR 27-AUG-1999; 990S-0151080.  
PR 30-AUG-1999; 990S-0151303.  
PR 31-AUG-1999; 990S-0151338.  
PR 01-SEP-1999; 990S-0151930.  
PR 07-SEP-1999; 990S-0152363.  
PR 10-SEP-1999; 990S-0153070.  
PR 13-SEP-1999; 990S-0153758.  
PR 15-SEP-1999; 990S-0154018.  
PR 16-SEP-1999; 990S-0154039.  
PR 20-SEP-1999; 990S-0154779.  
PR 22-SEP-1999; 990S-0155139.  
PR 23-SEP-1999; 990S-0155486.  
PR 24-SEP-1999; 990S-0155659.  
PR 28-SEP-1999; 990S-0158458.  
PR 29-SEP-1999; 990S-0158596.  
PR 04-OCT-1999; 990S-0157117.  
PR 05-OCT-1999; 990S-0157753.  
PR 06-OCT-1999; 990S-0157865.  
PR 07-OCT-1999; 990S-0158029.  
PR 08-OCT-1999; 990S-0158232.  
PR 12-OCT-1999; 990S-0158369.  
PR 13-OCT-1999; 990S-0159293.  
PR 13-OCT-1999; 990S-0159294.  
PR 13-OCT-1999; 990S-0159295.  
PR 14-OCT-1999; 990S-0159329.  
PR 14-OCT-1999; 990S-0159330.  
PR 14-OCT-1999; 990S-0159331.  
PR 14-OCT-1999; 990S-0159637.  
PR 18-OCT-1999; 990S-0159638.  
PR 18-OCT-1999; 990S-0159584.  
PR 21-OCT-1999; 990S-0160741.  
PR 21-OCT-1999; 990S-0160767.  
PR 21-OCT-1999; 990S-0160768.  
PR 21-OCT-1999; 990S-0160770.  
PR 21-OCT-1999; 990S-0160814.  
PR 21-OCT-1999; 990S-0160815.  
PR 22-OCT-1999; 990S-0160980.  
PR 22-OCT-1999; 990S-0160981.  
PR 22-OCT-1999; 990S-0160989.  
PR 25-OCT-1999; 990S-0161404.  
PR 25-OCT-1999; 990S-0161405.  
PR 25-OCT-1999; 990S-0161406.  
PR 26-OCT-1999; 990S-0161359.  
PR 26-OCT-1999; 990S-0161360.  
PR 26-OCT-1999; 990S-0161361.

PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 3.9%; Score 7; DB 21; Length 334;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 LAYDKAA 101  
 |||||  
 DB 189 LAYDKAA 195

RESULT 34  
 AAG71496  
 ID AAG71496 standard; Protein; 338 AA.

XX AAG71496;

XX 31-JUL-2001 (first entry)

DE Human olfactory receptor polypeptide, SEQ ID NO: 1177.

XX Human; olfactory receptor; OR; primary scent determination;  
 KW secondary scent determination; polypeptide library; odour receptor;  
 KW scent profile; scent fingerprint; scent representation.

XX Homo sapiens.

XX WO200127158-A2.

XX 19-APR-2001.

XX 06-OCT-2000; 2000WO-US27582.

XX 08-OCT-1999; 99US-0158615.

XX 24-FEB-2000; 2000US-0184809.

XX (DIGI-) DIGISCENTS.

PA (YEDA ) YEDA RES & DEV CO LTD.

XX Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

XX WPI; 2001-290713/30.

XX New polynucleotides which encode polypeptides involved in olfactory  
 PT sensation for identifying olfactory agonists and antagonists -

XX Claim 11; Page 693-694; 1857pp; English.

CC The present sequence is an olfactory receptor which is encoded by  
 CC one of a number of novel polynucleotides. The polynucleotides can be  
 CC used in screening for olfactory agonists and antagonists. The methods  
 CC allow for the determination of primary scents and the identification  
 CC of the odour receptors used to detect these primary scents. The methods  
 CC also enable determination of secondary scents and the identification of  
 CC combinations of odour receptors that are involved in detecting such  
 CC secondary scents. This enables the construction of a scent representation  
 CC (also called a scent fingerprint or scent profile), which may be used to  
 CC re-create and edit scents. Libraries of olfactory receptors are useful  
 CC for determining the interaction pattern of a composition with the  
 CC receptors, and can be used for determining differences in the olfactory  
 CC faculties of different individuals.

SQ Sequence 338 AA;

Query Match 3.9%; Score 7; DB 22; Length 338;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SLTLKLP 20  
 |||||

DB 167 SLTLKLP 173

RESULT 35

AAG72390  
 ID AAG72390 standard; Protein; 338 AA.

XX AAG72390;

XX 31-JUL-2001 (first entry)

XX Human OR-like polypeptide query sequence, SEQ ID NO: 2071.

XX Human; olfactory receptor; OR; primary scent determination;  
 KW secondary scent determination; polypeptide library; odour receptor;  
 KW scent profile; scent fingerprint; scent representation.

XX Homo sapiens.

XX WO200127158-A2.

XX 19-APR-2001.

XX 06-OCT-2000; 2000WO-US27582.

XX 08-OCT-1999; 99US-0158615.

XX 24-FEB-2000; 2000US-0184809.

XX (DIGI-) DIGISCENTS.

PA (YEDA ) YEDA RES & DEV CO LTD.

XX Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

XX WPI; 2001-290713/30.

XX New polynucleotides which encode polypeptides involved in olfactory  
 PT sensation for identifying olfactory agonists and antagonists -

XX Example 6; Page 1388-1389; 1857pp; English.

CC The present sequence is a polypeptide encoded by one of 344 newly mined  
 CC human genes. It was used as a query sequence in a database search of  
 CC olfactory receptor (OR)-like sequences. The invention relates to isolated  
 CC polynucleotides encoding polypeptides involved in olfactory sensation.  
 CC The polynucleotides can be used in screening for olfactory agonists and  
 CC antagonists. The methods allow for the determination of primary scents  
 CC and the identification of the odour receptors used to detect these  
 CC primary scents. The methods also enable determination of secondary scents  
 CC and the identification of combinations of odour receptors that are  
 CC involved in detecting such secondary scents. This enables the  
 CC construction of a scent representation (also called a scent fingerprint  
 CC or scent profile), which may be used to re-create and edit scents.  
 CC Libraries of olfactory receptors are useful for determining the  
 CC interaction pattern of a composition with the receptors, and can be  
 CC used for determining differences in the olfactory faculties of different  
 CC individuals.

SQ Sequence 338 AA;

Query Match 3.9%; Score 7; DB 22; Length 338;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SLTLKLP 20  
 |||||

DB 167 SLTLKLP 173

RESULT 36

AAW17820  
 ID AAW17820 standard; Protein; 466 AA.

XX AAW17820;



XX 29-JUL-1997 (first entry)  
 XX Pseudomonas putida amidase.  
 DE Amidase; nitrile hydratase; stereospecific enzyme; chiral compound;  
 KW amide; carboxylic acid.  
 XX Pseudomonas putida strain NRRL 18668.  
 OS WO9712964-A2.  
 PN 10-APR-1997.  
 PD 03-OCT-1996; 96WO-US15969.  
 XX 06-OCT-1995; 95US-0004914.  
 PR (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX Fallon RD, Nelson MJ, Payne MS;  
 PI WPI: 1997-226208/20.  
 DR N-PSDB; AAT66442.  
 XX Nucleic acid encoding stereospecific nitrile hydratase and amidase -  
 PT also transformed cells expressing these enzymes, useful for  
 PT conversion of racemic nitrile compounds to chiral amide(s) and acids  
 XX Claim 33; Page 58-60; 85pp; English.  
 PS The amidase (AAW17820) of Pseudomonas putida NRRL 18668 is capable of  
 CC converting a racemic mixture of C3 to C6 amides to the corresponding  
 CC enantiomeric R- or S-carboxylic acids. Its amino acid sequence was  
 CC deduced from a gene sequence (AAT66442) identified upstream of the  
 CC 18668 nitrile hydratase gene (AAT66440). The 18668 amidase is  
 CC distinct from the amidase isolated from P. putida B23 (FERM B-187).  
 CC Co-expression of the 18668 amidase with nitrile hydratase alpha and  
 CC beta subunits (see also AAW17816-17) in microbial host cells allows  
 CC the conversion of racemic nitrile compounds to the corresponding  
 CC enantiomeric R- or S-carboxylic acids.  
 XX Sequence 466 AA;  
 SQ  
 Query Match 3.9%; Score 7; DB 18; Length 466;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 91 LAGLAY 97  
 DB 443 LAGLAY 449  
 RESULT 37  
 AAW55454  
 ID AAW55454 standard; Protein; 511 AA.  
 XX AC AAW55454;  
 XX 24-JUN-1998 (first entry)  
 XX H. pylori ORF 02gel1622\_875260\_f3\_36 cell envelope OMP.  
 DE Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacteria; life cycle; activator;  
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.  
 XX Helicobacter pylori.  
 OS WO9737044-A1.  
 PN 09-OCT-1997.  
 PD  
 XX

PF 27-MAR-1997; 97WO-US05223.  
 XX 06-DEC-1996; 96US-0761318.  
 PR 29-MAR-1996; 96US-0625811.  
 PR 02-APR-1996; 96US-0758731.  
 PR 25-OCT-1996; 96US-0736905.  
 XX 28-OCT-1996; 96US-0738859.  
 XX (ASTR ) ASTRA AB.  
 XX Alm RA, Smith D;  
 PI WPI: 1997-503122/46.  
 DR N-PSDB; AAV24863.  
 XX Helicobacter pylori nucleic acid sequences and encoded  
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
 PT infection and for diagnosis of H. pylori infection  
 XX Claims 14,80; Page 661-662; 1145pp; English.  
 XX This sequence is a H. pylori cell envelope outer membrane protein  
 CC (OMP) having a terminal Phe residue.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors. The  
 CC DNA and probes derived from it may be used for the identification of  
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic  
 CC acid sequences complementary to the DNA act as antisense sequences and  
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies  
 CC against the protein can be used in immunoassays to evaluate the abundance  
 CC and distribution of H. pylori-specific antigens. The genomic sequence of  
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated  
 CC by mechanically shearing the bacterial DNA. The sequences were analysed  
 CC for ORF of at least 180 nucleotides, and the predicted coding regions  
 CC defined by computer evaluation. To identify likely H. pylori antigens for  
 CC vaccine development, the amino acid sequences predicted from various ORF  
 CC were analysed for significant homology to other known or expected  
 CC membrane proteins. Having identified and determined the sequences of  
 CC interest, particular regions can be isolated from H. pylori by PCR  
 CC amplification for recombinant polypeptide production, e.g. in E. coli  
 CC hosts.  
 XX Sequence 511 AA;  
 SQ  
 Query Match 3.9%; Score 7; DB 18; Length 511;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 12 KGSUTLK 18  
 DB 283 KGSUTLK 289  
 RESULT 38  
 AAW98255  
 ID AAW98255 standard; Protein; 511 AA.  
 XX AC AAW98255;  
 XX 31-MAR-1999 (first entry)  
 DE H. pylori GHPO 1559 protein.  
 XX GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;  
 KW peptic ulcer disease.  
 XX Helicobacter pylori.  
 OS WO9843478-A1.  
 PN 08-OCT-1998.  
 PD  
 XX

PF 01-APR-1998; 98WO-US06371.  
 XX 29-JUL-1997; 97US-0902615.  
 PR 01-APR-1997; 97US-0833457.  
 PR 24-JUN-1997; 97US-0881227.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
 XX Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;  
 PI WPI; 1998-542293/46.  
 DR N-PSDB; AAX13974.  
 XX New isolated Helicobacter polynucleotides - used to develop products  
 PT for the diagnosis, prevention and treatment of Helicobacter  
 PT infections and gastrointestinal diseases  
 XX Claim 8; Page 241-244; 2054pp; English.  
 XX This sequence represents a Helicobacter pylori GHPO protein of the  
 CC invention. The polypeptides can be used for preventing or treating  
 CC Helicobacter infections, and gastroduodenal diseases associated with  
 CC these infections, including acute, chronic, and atrophic gastritis, and  
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be  
 CC used for the production of antibodies. The products can also be used for  
 CC detection and diagnosis.  
 XX SQ Sequence 511 AA;

Query Match 3.9%; Score 7; DB 19; Length 511;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 KGSUTLK 18  
 |||||  
 Db 283 KGSUTLK 289

RESULT 39  
 AAW71519  
 ID AAW71519 standard; Protein; 511 AA.  
 XX AC AAW71519;  
 XX 09-NOV-1998 (first entry)  
 XX Helicobacter polypeptide GHPO 1559.  
 XX GHPO 1559; Infection; therapy; diagnosis; vaccine; gastritis;  
 KW ulcer.  
 XX Helicobacter pylori.  
 XX WO9821225-A1.  
 XX 22-MAY-1998.  
 XX 14-NOV-1997; 97WO-US21353.  
 XX 29-JUL-1997; 97US-0902615.  
 PR 14-NOV-1996; 96US-0749031.  
 PR 01-APR-1997; 97US-0831309.  
 PR 01-APR-1997; 97US-0833457.  
 PR 01-APR-1997; 97US-0834705.  
 PR 24-JUN-1997; 97US-0881227.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
 PA (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
 XX Al-Garawi A, Haas R, Kleanthous H, Meyer T, Miller C;  
 PI Odenbreit S, Tomb J;

XX WPI; 1998-297855/26.  
 DR N-PSDB; AAV52054.  
 XX Helicobacter polynucleotide and polypeptide sequences - useful to  
 PT treat or prevent gastrointestinal infection  
 XX Claim 1; Page 208-209; 362pp; English.  
 XX This claimed Helicobacter pylori polypeptide, designated GHPO 1559,  
 CC can be used in vaccination methods for preventing or treating  
 CC Helicobacter infection. 85 Helicobacter polypeptides (see  
 CC AAW71474-W71558) are claimed, as well as isolated polynucleotides  
 CC (see AAV52009-93) that encode them. The invention also provides:  
 CC methods for producing these Helicobacter polypeptides in  
 CC recombinant host systems, and related expression cassettes, vectors  
 CC and transformed or transfected host cells; live vaccine vectors  
 CC that contain the polynucleotides of the invention and which can be  
 CC used to prevent or treat Helicobacter infection; therapeutic and/or  
 CC prophylactic methods involving administration of polynucleotide  
 CC molecules, polypeptides or monospecific antibodies; methods for  
 CC detecting the presence of Helicobacter in samples using e.g.  
 CC the polypeptides or monospecific antibodies; and methods for  
 CC purifying the polypeptides by antibody-based affinity  
 CC chromatography.  
 XX SQ Sequence 511 AA;

Query Match 3.9%; Score 7; DB 19; Length 511;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 KGSUTLK 18  
 |||||  
 Db 283 KGSUTLK 289

RESULT 40  
 AAY17176  
 ID AAY17176 standard; Protein; 511 AA.  
 XX AC AAY17176;  
 XX 03-AUG-1999 (first entry)  
 XX H. pylori outer membrane polypeptide.  
 XX Outer membrane polypeptide; OMP; vaccine; H. pylori infection; humoral;  
 KW cellular immune response.  
 XX Helicobacter pylori.  
 XX WO9921959-A2.  
 XX 06-MAY-1999.  
 XX 28-OCT-1998; 98WO-US22883.  
 PR 17-DEC-1997; 97US-0993001.  
 PR 28-OCT-1997; 97US-0959131.  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 PA Alm RA, Ellis RW, Guld BC, Noonan BM, Smith D;  
 XX WPI; 1999-326698/27.  
 DR N-PSDB; AAX75795.  
 XX Cellular vaccine against Helicobacter pylori  
 XX Claim 7; Page 229-231; 352pp; English.  
 XX The invention relates to a vaccine for preventing or treating infections

CC by Helicobacter pylori. The vaccine contains at least one isolated  
CC H. pylori polypeptide, or its fragments, in a carrier, where the  
CC carrier is a Salmonella, Vibrio cholerae or Shigella vector containing a  
CC nucleic acid encoding the H. pylori polypeptide. The vaccines induce  
CC humoral and cellular immune responses. The vaccines are used to treat or  
CC prevent infections by H. pylori. Sequences AAX75779 to AAX75837 represent  
CC nucleic acid sequences encoding H. pylori outer membrane polypeptides  
CC (OMPs) AAY17160 to AAY17218.  
XX  
SQ Sequence 511 AA;

Query Match 3.9%; Score 7; DB 20; Length 511;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 KGSLTLK 18  
          |||||  
DB 283 KGSLTLK 289

Search completed: May 12, 2003, 09:47:38  
Job time : 61 secs



;  
; PRIOR APPLICATION NUMBER: 60/056,878  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,662  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,872  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,882  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,637  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,903  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,888  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,879  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,880  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,894  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,911  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,636  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,874  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,910  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,864  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,631  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,845  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,892  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/047,595  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/057,761  
; PRIOR FILING DATE: 05-Sep-1997  
; PRIOR APPLICATION NUMBER: 60/047,599  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/047,588  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/047,585  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/047,586  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/047,590  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/047,594  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/047,589  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/047,593  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/047,614  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/043,578  
; PRIOR FILING DATE: 1997-04-11  
; PRIOR APPLICATION NUMBER: 60/043,576  
; PRIOR FILING DATE: 1997-04-11  
; PRIOR APPLICATION NUMBER: 60/047,501  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/043,670  
; PRIOR FILING DATE: 1997-04-11  
; PRIOR APPLICATION NUMBER: 60/056,632  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,664  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,876  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,881

;  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,909  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,875  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,862  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,887  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,908  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/048,964  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/057,650  
; PRIOR FILING DATE: 1997-09-05  
; PRIOR APPLICATION NUMBER: 60/056,884  
; PRIOR FILING DATE: 1997-08-22  
; NUMBER OF SEQ ID NOS: 280  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 184  
; LENGTH: 168

Query Match 3.4% Score 6; DB 9; Length 168;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 LTLKLP 20  
| | | | |  
Db 43 LTLKLP 48

Search completed: May 12, 2003, 09:56:07  
Job time : 41 secs



```
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/983,802
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/227,357
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/US98/13684
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,926
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,793
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,925
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,929
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,803
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,732
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,931
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,932
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,916
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,930
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,918
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,920
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,733
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,795
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,919
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,928
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,722
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,723
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,948
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,949
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,953
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,950
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,947
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,964
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/056,360
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,684
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,984
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,954
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,785
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,664
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,660
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,661
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12

; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 196
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (61)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (142)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (149)
; OTHER INFORMATION: Xaa equals stop translation
US-09-983-802-196

Query Match 3.4%; Score 6; DB 9; Length 149;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 QDGSY 62
Db 43 QDGSY 48
|||||

RESULT 38
US-09-925-300-1688
; Sequence 1688, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1688
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-300-1688

Query Match 3.4%; Score 6; DB 10; Length 153;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 139 LRGKKK 144
Db 39 LRGKKK 44
|||||

RESULT 39
US-08-781-986A-5220
; Sequence 5220, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
```

; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOL, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4318  
; LENGTH: 132  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4318

Query Match 3.4%; Score 6; DB 9; Length 132;  
Best Local Similarity 100.0%; Pred. No. 3e+02; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0

QY 103 KTHLSA 108  
Db 123 KTHLSA 128  
|||||

RESULT 34  
US-09-862-540-51  
; Sequence 51, Application US/09862540  
; Publication No. US20030023992A1  
; GENERAL INFORMATION:  
; APPLICANT: Vogell, Gabriel  
; TITLE OF INVENTION: No. US20030023992A1e1 G Protein-Coupled Receptors  
; FILE REFERENCE: 002520S1  
; CURRENT APPLICATION NUMBER: US/09/862,540  
; CURRENT FILING DATE: 2001-05-22  
; PRIOR APPLICATION NUMBER: 60/206,138  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: 60/206,139  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: 60/208,976  
; PRIOR FILING DATE: 2000-06-02  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 51  
; LENGTH: 137  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-862-540-51

Query Match 3.4%; Score 6; DB 9; Length 137;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 LYGLNL 175  
Db 69 LYGLNL 74  
|||||

RESULT 35  
US-10-102-806-799  
; Sequence 799, Application US/10102806  
; Publication No. US20030054421A1  
; GENERAL INFORMATION:

; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA103P1C1  
; CURRENT APPLICATION NUMBER: US/10/102,806  
; CURRENT FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: 09/925,298  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05881  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 846  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 799  
; LENGTH: 142  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-102-806-799

Query Match 3.4%; Score 6; DB 9; Length 142;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 LRKKKK 144  
Db 28 LRKKKK 33  
|||||

RESULT 36  
US-10-125-258-10  
; Sequence 10, Application US/10125258  
; Publication No. US20030028920A1  
; GENERAL INFORMATION:  
; APPLICANT: Altier, Daniel J.  
; APPLICANT: Herrmann, Rafael  
; APPLICANT: Lu, Albert L.  
; APPLICANT: McCutchen, Billy F.  
; APPLICANT: Presnail, James K.  
; APPLICANT: Weaver, Janine L.  
; APPLICANT: Wong, James F. H.  
; TITLE OF INVENTION: Antimicrobial Polypeptides and Their  
; FILE REFERENCE: 35718/246215  
; CURRENT APPLICATION NUMBER: US/10/125,258  
; CURRENT FILING DATE: 2002-04-18  
; PRIOR APPLICATION NUMBER: 60/285,355  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 127  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 148  
; TYPE: PRT  
; ORGANISM: Manduca sexta  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: 141  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-10-125-258-10

Query Match 3.4%; Score 6; DB 9; Length 148;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 AGLAYD 98  
Db 86 AGLAYD 91  
|||||

RESULT 37  
US-09-983-802-196  
; Sequence 196, Application US/09983802  
; Publication No. US20030022185A1  
; GENERAL INFORMATION:



; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 41522  
; LENGTH: 104  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL161747.1  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.5  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.8  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.7  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.5  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.5  
; OTHER INFORMATION: EST\_HUMAN HIT: BF340331.1, EVALUE 2.00e-24  
; OTHER INFORMATION: SWISSPROT HIT: P53236, EVALUE 1.50e+00  
US-09-864-761-41522

Query Match 3.4%; Score 6; DB 10; Length 104;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 SLTLKL 19  
|||||  
Db 19 SLTLKL 24

RESULT 30  
US-10-092-154-844  
; Sequence 844, Application US/10092154  
; Publication No. US20030054375A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC009C1  
; CURRENT APPLICATION NUMBER: US/10/092,154  
; CURRENT FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 2003  
; Prior Application removed - See File Wrapper or Palm  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 844  
; LENGTH: 110  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (54)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-10-092-154-844

Query Match 3.4%; Score 6; DB 9; Length 110;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 NLSVDV 134  
|||||  
Db 69 NLSVDV 74

RESULT 31  
US-09-764-847-844

; Sequence 844, Application US/09764847  
; Patent No. US20020132767A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC009  
; CURRENT APPLICATION NUMBER: US/09/764,847  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 2003  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 844  
; LENGTH: 110  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (54)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-847-844

Query Match 3.4%; Score 6; DB 10; Length 110;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 NLSVDV 134  
|||||  
Db 69 NLSVDV 74

RESULT 32  
US-10-125-258-8  
; Sequence 8, Application US/10125258  
; Publication No. US20030028920A1  
; GENERAL INFORMATION:  
; APPLICANT: Altier, Daniel J.  
; APPLICANT: Herrmann, Rafael  
; APPLICANT: Lu, Albert L.  
; APPLICANT: McCutchen, Billy F.  
; APPLICANT: Presnail, James K.  
; APPLICANT: Weaver, Janine L.  
; APPLICANT: Wong, James F. H.  
; TITLE OF INVENTION: Antimicrobial Polypeptides and Their  
; FILE REFERENCE: 35718/246215  
; CURRENT APPLICATION NUMBER: US/10/125,258  
; CURRENT FILING DATE: 2002-04-18  
; PRIOR APPLICATION NUMBER: 60/285,355  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 127  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 111  
; TYPE: PRT  
; ORGANISM: Manduca sexta  
US-10-125-258-8

Query Match 3.4%; Score 6; DB 9; Length 111;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 93 AGLAYD 98  
|||||  
Db 88 AGLAYD 93

RESULT 33  
US-09-738-626-4318  
; Sequence 4318, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI

; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 43033  
; LENGTH: 91  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL049575.7  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.8  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.72  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4  
; OTHER INFORMATION: SWISSPROT HIT: P45190, EVALUE 3.10e+00  
; OTHER INFORMATION: EST\_HUMAN HIT: AW904936.1, EVALUE 2.00e-31  
US-09-864-761-43033

Query Match 3.4%; Score 6; DB 10; Length 91;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 KGLLLV 158  
|||||  
Db 31 KGLLLV 36

RESULT 27  
US-09-867-550-612  
; Sequence 612, Application US/09867550  
; Patent No. US20020082206A1  
; GENERAL INFORMATION:  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Mehraban, Fuad,  
; APPLICANT: Conley, Pamela  
; APPLICANT: Law, Debbie  
; APPLICANT: Topper, James  
; TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells and  
; FILE REFERENCE: 21402-013 (Cura-313)  
; CURRENT APPLICATION NUMBER: US/09/867,550  
; CURRENT FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: USSN 60/208,427  
; PRIOR FILING DATE: 2000-05-30  
; NUMBER OF SEQ ID NOS: 2125  
; SOFTWARE: FastSeq for Windows version 4.0  
; SEQ ID NO 612  
; LENGTH: 93  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)  
; OTHER INFORMATION: wherein Xaa may be any one of Ala or Arg or Glu or Gln or Gly or  
; OTHER INFORMATION: Leu or Lys or Pro or Ser or Thr or Val  
US-09-867-550-612

Query Match 3.4%; Score 6; DB 10; Length 93;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 LLLVEA 160  
|||||  
Db 26 LLLVEA 31

RESULT 28

US-10-300-616-49  
; Sequence 49, Application US/10300616  
; Publication No. US20030082801A1  
; GENERAL INFORMATION:  
; APPLICANT: BARNES, ASHLEY A.  
; APPLICANT: WISE, ALAN  
; APPLICANT: MARSHALL, FIONA H.  
; APPLICANT: FRASER, NEIL J.  
; APPLICANT: WHITE, JULIE H. M.  
; APPLICANT: FOORD, STEVEN M.  
; TITLE OF INVENTION: NOVEL RECEPTOR  
; FILE REFERENCE: PG3558US2  
; CURRENT APPLICATION NUMBER: US/10/300,616  
; CURRENT FILING DATE: 2002-11-20  
; PRIOR APPLICATION NUMBER: GB9819420.2  
; PRIOR FILING DATE: 1998-09-07  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 49  
; LENGTH: 102  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-10-300-616-49

Query Match 3.4%; Score 6; DB 9; Length 102;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 FTKKEE 70  
|||||  
Db 21 FTKKEE 26

RESULT 29  
US-09-864-761-41522  
; Sequence 41522, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aeonics-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30

APPLICANT: Watson, James D.  
APPLICANT: Strachan, Lorna  
APPLICANT: Sleeman, Matthew  
APPLICANT: Onrust, Rene  
APPLICANT: Murison, James G.  
APPLICANT: Kumble, Krishanand D.  
TITLE OF INVENTION: Compositions Isolated From Skin Cells  
TITLE OF INVENTION: and Methods for Their Use  
FILE REFERENCE: 11000.101c4u  
CURRENT APPLICATION NUMBER: US/09/866.050A  
CURRENT FILING DATE: 2001-05-24  
NUMBER OF SEQ ID NOS: 725  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 134  
LENGTH: 90  
TYPE: PRT  
ORGANISM: Rat  
US-09-866-050A-134

Query Match 3.4%; Score 6; DB 9; Length 90;  
Best Local Similarity 100.0%; Pred. No. 2.le+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 KAASKT 104  
|||||  
Db 31 KAASKT 36

## RESULT 25

US-09-864-761-43578  
Sequence 43578, Application US/09864761  
Patent No. US20020048763A1

## GENERAL INFORMATION:

APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

FILE REFERENCE: Aecomica-X-1

CURRENT APPLICATION NUMBER: US/09/864, 761

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 43578  
LENGTH: 90  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC004147.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3  
OTHER INFORMATION: SWISSPROT HIT: Q62962, EVALUATE 6.00e-35  
OTHER INFORMATION: EST\_HUMAN HIT: A1277385.1, EVALUATE 2.00e-16  
US-09-864-761-43578

Query Match 3.4%; Score 6; DB 10; Length 90;  
Best Local Similarity 100.0%; Pred. No. 2.le+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 1; Gaps 0;

QY 154 GLLLVE 159  
|||||  
Db 56 GLLLVE 61

## RESULT 26

US-09-864-761-43033  
Sequence 43033, Application US/09864761  
Patent No. US20020048763A1

## GENERAL INFORMATION:

APPLICANT: Penn, Sharron G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

FILE REFERENCE: Aecomica-X-1

CURRENT APPLICATION NUMBER: US/09/864, 761

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

US-10-079-854-110  
; Sequence 110, Application US/10079854  
; Publication No. US20030054368A1  
; GENERAL INFORMATION:

; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA121C1  
; CURRENT APPLICATION NUMBER: US/10/079,854  
; CURRENT FILING DATE: 2002-02-22  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 428  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 110  
; LENGTH: 76

; TYPE: PRT  
; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (17)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: misc\_feature

; LOCATION: (66)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-10-079-854-110

Query Match 3.4%; Score 6; DB 9; Length 76;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 LTLKLP 20

|||||

Db 53 LTLKLP 58

RESULT 21

US-09-764-878-110

; Sequence 110, Application US/09764878

; Patent No. US20020090615A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PA121

; CURRENT APPLICATION NUMBER: US/09/764,878

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 428

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 110

; LENGTH: 76

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (17)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: SITE

; LOCATION: (66)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-764-878-110

Query Match

Best Local Similarity 3.4%; Score 6; DB 10; Length 76;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 LTLKLP 20

|||||

Db 53 LTLKLP 58

RESULT 22

US-09-893-737-12

; Sequence 12, Application US/09893737

; Patent No. US2002011085A1

; GENERAL INFORMATION:

; APPLICANT: Sheppard, Paul O.

; APPLICANT: Presnell, Scott R.

; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS

; FILE REFERENCE: 00-41

; CURRENT APPLICATION NUMBER: US/09/893,737

; CURRENT FILING DATE: 2001-06-28

; PRIOR APPLICATION NUMBER: US 60/215,446

; PRIOR FILING DATE: 2000-06-30

; NUMBER OF SEQ ID NOS: 329

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 12

; LENGTH: 80

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-893-737-12

Query Match

Best Local Similarity 3.4%; Score 6; DB 10; Length 80;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 KGLLLV 158

|||||

Db 5 KGLLLV 10

RESULT 23

US-09-738-626-6220

; Sequence 6220, Application US/09738626

; Publication No. US20020197605A1

; GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI

; APPLICANT: MIZOGUCHI, HIROSHI

; APPLICANT: ANDO, SEIKO

; APPLICANT: HAYASHI, MIKIO

; APPLICANT: OCHIAI, KEIKO

; APPLICANT: YOKOI, HARUHIKO

; APPLICANT: TATEISHI, NAKO

; APPLICANT: SENOH, AKIHIRO

; APPLICANT: IKEDA, MASATO

; APPLICANT: OZAKI, AKIO

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-125

; CURRENT APPLICATION NUMBER: US/09/738,626

; CURRENT FILING DATE: 2000-12-18

; PRIOR APPLICATION NUMBER: JP 99/377484

; PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: JP 00/159162

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: JP 00/280988

; PRIOR FILING DATE: 2000-08-03

; NUMBER OF SEQ ID NOS: 7059

; SOFTWARE: PatentIn ver. 3.0

; SEQ ID NO 6220

; LENGTH: 82

; TYPE: PRT

; ORGANISM: Corynebacterium glutamicum

US-09-738-626-6220

Query Match

Best Local Similarity 3.4%; Score 6; DB 9; Length 82;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 DALTLR 92

|||||

Db 59 DALTLR 64

RESULT 24

US-09-866-050A-134

; Sequence 134, Application US/09866050A

; Publication No. US20030040471A1

; GENERAL INFORMATION:

```
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO ALL136527.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.2
; OTHER INFORMATION: SWISSPROT HIT: P07954, EVALUE 1.00e-29
; OTHER INFORMATION: EST_HUMAN HIT: BE293788.1, EVALUE 1.00e-28
US-09-864-761-41976

Query Match          3.4%; Score 6; DB 10; Length 68;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  96 AYDKAA 101
    |||||
Db  32 AYDKAA 37

RESULT 18
US-09-864-761-46121
; Sequence 46121, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aesomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
```

```
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 46121
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010618.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.9
; OTHER INFORMATION: EST_HUMAN HIT: AW630291.1, EVALUE 9.00e-34
; OTHER INFORMATION: SWISSPROT HIT: Q03018, EVALUE 5.10e-01
US-09-864-761-46121
```

```
Query Match          3.4%; Score 6; DB 10; Length 68;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy  157 LVEADY 162
    |||||
Db  22 LVEADY 27
```

## RESULT 19

```
US-09-925-300-1548
; Sequence 1548, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1548
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (37)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (59)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (63)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1548
```

```
Query Match          3.4%; Score 6; DB 10; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy  102 SKTHLS 107
    |||||
Db  18 SKTHLS 23
```

## RESULT 20

;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 34896  
;; LENGTH: 64  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC010732.1  
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.1  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 9.5  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9  
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.3  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6  
;; OTHER INFORMATION: EST\_HUMAN HIT: BE541170.1, EVALUATE 8.80e-01  
US-09-864-761-34896

Query Match 3.4%; Score 6; DB 10; Length 64;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SUTLKL 19  
DB 31 SUTLKL 36

RESULT 16  
US-09-867-550-904  
;; Sequence 904, Application US/09867550  
;; Patent No. US20020082206A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Leach, Martin D.  
;; APPLICANT: Mehrahan, Fuad,  
;; APPLICANT: Conley, Pamela  
;; APPLICANT: Law, Debbie  
;; APPLICANT: Topper, James  
;; TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells and  
;; FILE REFERENCE: 21402-013 (Cura-313)  
;; CURRENT APPLICATION NUMBER: US/09/867,550  
;; CURRENT FILING DATE: 2001-09-20

;; PRIOR APPLICATION NUMBER: USSN 60/208,427  
;; PRIOR FILING DATE: 2000-05-30  
;; NUMBER OF SEQ ID NOS: 2125  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 904  
;; LENGTH: 64  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-867-550-904

Query Match 3.4%; Score 6; DB 10; Length 64;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 SKTHLS 107  
DB 23 SKTHLS 28

RESULT 17  
US-09-864-761-41976  
;; Sequence 41976, Application US/09864761  
;; Patent No. US20020048763A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Penn, Sharon G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
;; FILE REFERENCE: Aeonica-X-1  
;; CURRENT APPLICATION NUMBER: US/09/864,761  
;; CURRENT FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 41976  
;; LENGTH: 68

```
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40107
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005317.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
; OTHER INFORMATION: EST_HUMAN HIT: BE615382.1, EVALUE 7.40e+00
; OTHER INFORMATION: SWISSPROT HIT: P07646, EVALUE 2.80e+00
US-09-864-761-40107

Query Match 3.4%; Score 6; DB 10; Length 37;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 LTLRAG 94
Db 11 LTLRAG 16
|||||

RESULT 13
US-10-122-822-25
; Sequence 25, Application US/10122822
; Publication No. US20030084477A1
; GENERAL INFORMATION:
; APPLICANT: Purdue Research Foundation
; TITLE OF INVENTION: GENES ENHANCING DISEASE RESISTANCE IN PLANTS
; FILE REFERENCE: 7024-511
; CURRENT APPLICATION NUMBER: US/10/122,822
; CURRENT FILING DATE: 2002-04-15
; PRIOR APPLICATION NUMBER: US 09/202,161
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: PCT/US97/10382
; PRIOR FILING DATE: 1997-06-12
; PRIOR APPLICATION NUMBER: 60/046,494
; PRIOR FILING DATE: 1997-05-14
; PRIOR APPLICATION NUMBER: 60/019,633
; PRIOR FILING DATE: 1996-06-12
```

```
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: ASCII
; SEQ ID NO 25
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Lycopersicon esculentum
US-10-122-822-25

Query Match 3.4%; Score 6; DB 9; Length 59;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 96 AYDKAA 101
Db 41 AYDKAA 46
|||||

RESULT 14
US-09-867-550-1856
; Sequence 1856, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells a
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1856
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-867-550-1856

Query Match 3.4%; Score 6; DB 10; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 154 GLLIVE 159
Db 38 GLLIVE 43
|||||

RESULT 15
US-09-864-761-34896
; Sequence 34896, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
```

;; PRIOR APPLICATION NUMBER: 60/070,923  
;; PRIOR FILING DATE: 1997-12-18  
;; PRIOR APPLICATION NUMBER: 60/068,007  
;; PRIOR FILING DATE: 1997-12-18  
;; PRIOR APPLICATION NUMBER: 60/068,057  
;; PRIOR FILING DATE: 1997-12-18  
;; PRIOR APPLICATION NUMBER: 60/068,006  
;; PRIOR FILING DATE: 1997-12-18  
;; PRIOR APPLICATION NUMBER: 60/068,369  
;; PRIOR FILING DATE: 1997-12-19  
;; PRIOR APPLICATION NUMBER: 60/068,367  
;; PRIOR FILING DATE: 1997-12-19  
;; PRIOR APPLICATION NUMBER: 60/068,368  
;; PRIOR FILING DATE: 1997-12-19  
;; PRIOR APPLICATION NUMBER: 60/068,169  
;; PRIOR FILING DATE: 1997-12-19  
;; PRIOR APPLICATION NUMBER: 60/068,053  
;; PRIOR FILING DATE: 1997-12-18  
;; PRIOR APPLICATION NUMBER: 60/068,064  
;; PRIOR FILING DATE: 1997-12-18  
;; PRIOR APPLICATION NUMBER: 60/068,054  
;; PRIOR FILING DATE: 1997-12-18  
;; PRIOR APPLICATION NUMBER: 60/068,008  
;; PRIOR FILING DATE: 1997-12-18  
;; PRIOR APPLICATION NUMBER: 60/068,365  
;; PRIOR FILING DATE: 1997-12-19  
;; NUMBER OF SEQ ID NOS: 672  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 537  
;; LENGTH: 24  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-097-065-537

Query Match 3.4%; Score 6; DB 9; Length 24;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 14 SLTLKL 19  
| | | | | |  
DB 12 SLTLKL 17

RESULT 11  
US-09-864-761-36829  
;; Sequence 36829, Application US/09864761  
;; Patent No. US2002048763A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Penn, Sharron G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; FILE REFERENCE: Aeonica-X-1  
;; CURRENT APPLICATION NUMBER: US/09/864,761  
;; PRIOR FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 36829  
;; LENGTH: 34  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC011174.2  
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.8  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2  
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9  
US-09-864-761-36829

Query Match 3.4%; Score 6; DB 10; Length 34;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 99 KAASKT 104  
| | | | | |  
DB 10 KAASKT 15

RESULT 12  
US-09-864-761-40107  
;; Sequence 40107, Application US/09864761  
;; Patent No. US2002048763A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Penn, Sharron G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL F  
;; FILE REFERENCE: Aeonica-X-1  
;; CURRENT APPLICATION NUMBER: US/09/864,761  
;; PRIOR FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27



```
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-881-752A-110

Query Match          3.9%; Score 7; DB 10; Length 511;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 KGSJTLK 18
Db 283 KGSJTLK 289

RESULT 7
US-09-738-626-3810
; Sequence 3810, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIRO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patentin ver. 3.0
; SEQ ID NO 3810
; LENGTH: 795
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3810

Query Match          3.9%; Score 7; DB 9; Length 795;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 LLLVEAD 161
Db 141 LLLVEAD 147

RESULT 8
US-10-024-918-10
; Sequence 10, Application US/10024918
; Patent No. US20020168718A1
; GENERAL INFORMATION:
; APPLICANT: Hubbell, Jeffrey
; APPLICANT: Schense, Jason
; APPLICANT: Zisch, Andreas
; APPLICANT: Hall, Heike
; TITLE OF INVENTION: ENZYME-MEDIATED MODIFICATION OF FIBRIN FOR TISSUE ENGINEERING
; FILE REFERENCE: CIT 2606 CIP
; CURRENT APPLICATION NUMBER: US/10/024,918
; CURRENT FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 10
; LENGTH: 8
```

```
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: random coupling polylysine peptide attached to active peptide
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: dansylated leucine
US-10-024-918-10

Query Match          3.4%; Score 6; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 3.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 LRKGGK 144
Db 1 LRKGGK 6

RESULT 9
US-10-150-111-181
; Sequence 181, Application US/10150111
; Publication No. US20030078386A1
; GENERAL INFORMATION:
; APPLICANT: Rubin et al.
; TITLE OF INVENTION: Secreted Protein HPEAD48
; FILE REFERENCE: P2018PDI
; CURRENT APPLICATION NUMBER: US/10/150,111
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 09/288,143
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: PCT/US98/21142
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/061,463
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,529
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/071,498
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,527
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,536
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,532
; PRIOR FILING DATE: 1997-10-09
; NUMBER OF SEQ ID NOS: 219
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 181
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-150-111-181

Query Match          3.4%; Score 6; DB 9; Length 21;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 LNDALT 90
Db 2 LNDALT 7

RESULT 10
US-10-097-065-537
; Sequence 537, Application US/10097065
; Publication No. US2003005236A1
; GENERAL INFORMATION:
; APPLICANT: Moore, Paul A. et al.
; TITLE OF INVENTION: 110 Human Secreted Proteins
; FILE REFERENCE: P2021P1
; CURRENT APPLICATION NUMBER: US/10/097,065
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: PCT/US98/27059
; PRIOR FILING DATE: 1998-12-17
```

APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 6630  
LENGTH: 93  
TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-6630

Query Match 3.9%; Score 7; DB 9; Length 93;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 IKGLLLV 158  
|||||||  
DB 82 IKGLLLV 88

## RESULT 4

US-10-278-173-150  
Sequence 150, Application US/10278173  
Publication No. US20030061637A1  
GENERAL INFORMATION:  
APPLICANT: Jiang, Cai-Zhong  
APPLICANT: Broun, Pierre  
APPLICANT: Riechmann, Jose-Luis  
APPLICANT: Pineda, Onaira  
APPLICANT: Zhang, James  
APPLICANT: Yu, Guo-Liang  
APPLICANT: Pilgrim, Marsha  
APPLICANT: Keddle, James  
APPLICANT: Heard, Jacqueline  
APPLICANT: Reuber, Lynne  
APPLICANT: Ratcliffe, Oliver  
APPLICANT: Adam, Luc  
APPLICANT: Samaha, Raymond  
TITLE OF INVENTION: POLYNUCLEOTIDES FOR ROOT TRAIT ALTERATION  
FILE REFERENCE: MBI-009  
CURRENT APPLICATION NUMBER: US/10/278,173  
CURRENT FILING DATE: 2002-10-21  
PRIOR APPLICATION NUMBER: US/09/533,392  
PRIOR FILING DATE: 2000-03-22  
PRIOR APPLICATION NUMBER: 60/125,814  
PRIOR FILING DATE: 1999-03-23  
NUMBER OF SEQ ID NOS: 177  
SOFTWARE: PatentIn ver. 2.1  
SEQ ID NO 150  
LENGTH: 218  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
OTHER INFORMATION: G26  
US-10-278-173-150

Query Match 3.9%; Score 7; DB 9; Length 218;  
Best Local Similarity 100.0%; Pred. No. 49;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 95 LAYDKAA 101  
|||||||  
DB 107 LAYDKAA 113

## RESULT 5

US-10-278-173-82  
Sequence 82, Application US/10278173  
Publication No. US20030061637A1  
GENERAL INFORMATION:  
APPLICANT: Jiang, Cai-Zhong  
APPLICANT: Broun, Pierre  
APPLICANT: Riechmann, Jose-Luis  
APPLICANT: Pineda, Onaira  
APPLICANT: Zhang, James  
APPLICANT: Yu, Guo-Liang  
APPLICANT: Pilgrim, Marsha  
APPLICANT: Keddle, James  
APPLICANT: Heard, Jacqueline  
APPLICANT: Reuber, Lynne  
APPLICANT: Ratcliffe, Oliver  
APPLICANT: Adam, Luc  
APPLICANT: Samaha, Raymond  
TITLE OF INVENTION: POLYNUCLEOTIDES FOR ROOT TRAIT ALTERATION  
FILE REFERENCE: MBI-009  
CURRENT APPLICATION NUMBER: US/10/278,173  
CURRENT FILING DATE: 2002-10-21  
PRIOR APPLICATION NUMBER: US/09/533,392  
PRIOR FILING DATE: 2000-03-22  
PRIOR APPLICATION NUMBER: 60/125,814  
PRIOR FILING DATE: 1999-03-23  
NUMBER OF SEQ ID NOS: 177  
SOFTWARE: PatentIn ver. 2.1  
SEQ ID NO 82  
LENGTH: 334  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
OTHER INFORMATION: G5  
US-10-278-173-82

Query Match 3.9%; Score 7; DB 9; Length 334;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 LAYDKAA 101  
|||||||  
DB 189 LAYDKAA 195

## RESULT 6

US-09-881-752A-110  
Sequence 110, Application US/09881752A  
Patent No. US20020115078A1  
GENERAL INFORMATION:  
APPLICANT: Kleanthous, Harold  
APPLICANT: Al-Garawi, Amal  
APPLICANT: Miller, Charles  
APPLICANT: Tomb, Jean-Francois  
APPLICANT: Oomen, Raymond P.  
TITLE OF INVENTION: Identification of Polynucleotides  
TITLE OF INVENTION: Encoding No. US20020115078A1el Helicobacter Polypeptides in  
TITLE OF INVENTION: Genome  
FILE REFERENCE: 06132/041002  
CURRENT APPLICATION NUMBER: US/09/881,752A  
CURRENT FILING DATE: 2001-05-15  
PRIOR APPLICATION NUMBER: US 08/833,457  
PRIOR FILING DATE: 1997-04-01  
NUMBER OF SEQ ID NOS: 370  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 110

969 6 3.4 620 9 US-10-127-901A-324 Sequence 324, App  
970 6 3.4 620 9 US-10-128-693A-324 Sequence 324, App  
971 6 3.4 620 9 US-10-131-813A-324 Sequence 324, App  
972 6 3.4 620 9 US-10-131-818A-324 Sequence 324, App  
973 6 3.4 620 9 US-10-131-823A-324 Sequence 324, App  
974 6 3.4 620 9 US-10-131-824A-324 Sequence 324, App  
975 6 3.4 620 9 US-10-131-830A-324 Sequence 324, App  
976 6 3.4 620 9 US-10-131-837A-324 Sequence 324, App  
977 6 3.4 620 9 US-10-137-872A-324 Sequence 324, App  
978 6 3.4 620 9 US-10-140-860-324 Sequence 324, App  
979 6 3.4 620 9 US-10-142-417-324 Sequence 324, App  
980 6 3.4 620 9 US-10-147-500-324 Sequence 324, App  
981 6 3.4 620 9 US-10-147-502-324 Sequence 324, App  
982 6 3.4 620 9 US-10-147-515-324 Sequence 324, App  
983 6 3.4 620 9 US-10-147-517-324 Sequence 324, App  
984 6 3.4 620 9 US-10-147-519-324 Sequence 324, App  
985 6 3.4 620 9 US-10-147-526-324 Sequence 324, App  
986 6 3.4 620 9 US-10-147-527-324 Sequence 324, App  
987 6 3.4 620 9 US-10-152-395-324 Sequence 324, App  
988 6 3.4 620 9 US-10-157-782-324 Sequence 324, App  
989 6 3.4 620 9 US-09-902-634-73 Sequence 73, Appl  
990 6 3.4 620 9 US-09-902-713-73 Sequence 73, Appl  
991 6 3.4 620 9 US-09-907-979-73 Sequence 73, Appl  
992 6 3.4 620 9 US-10-121-040-324 Sequence 324, App  
993 6 3.4 620 9 US-10-121-056-324 Sequence 324, App  
994 6 3.4 620 9 US-10-121-061-324 Sequence 324, App  
995 6 3.4 620 9 US-10-123-235-324 Sequence 324, App  
996 6 3.4 620 9 US-10-124-818-324 Sequence 324, App  
997 6 3.4 620 9 US-10-125-926A-324 Sequence 324, App  
998 6 3.4 620 9 US-10-125-930A-324 Sequence 324, App  
999 6 3.4 620 9 US-10-127-831A-324 Sequence 324, App  
1000 6 3.4 620 9 US-10-127-837A-324 Sequence 324, App

## ALIGNMENTS

RESULT 1  
US-09-996-140-11  
; Sequence 11, Application US/09996140  
; Patent No. US20020157136A1  
; GENERAL INFORMATION:  
; APPLICANT: Thomashow, Michael  
; APPLICANT: Stockinger, Eric  
; APPLICANT: Jaglo-Ottosen, Kirsten  
; APPLICANT: Gilmour, Sarah  
; APPLICANT: Zarka, Daniel  
; TITLE OF INVENTION: PLANT HAVING ALTERED ENVIRONMENTAL  
; FILE REFERENCE: MB10040US  
; CURRENT APPLICATION NUMBER: US/09/996,140  
; CURRENT FILING DATE: 2001-11-26  
; PRIOR APPLICATION NUMBER: 09/706,270  
; PRIOR FILING DATE: 1996-09-04  
; PRIOR APPLICATION NUMBER: 09/018,233  
; PRIOR FILING DATE: 1998-02-03  
; PRIOR APPLICATION NUMBER: 09/018,235  
; PRIOR FILING DATE: 1998-02-03  
; PRIOR APPLICATION NUMBER: 09/018,227  
; PRIOR FILING DATE: 1998-02-03  
; PRIOR APPLICATION NUMBER: 09/018,234  
; PRIOR FILING DATE: 1998-02-03  
; PRIOR APPLICATION NUMBER: 09/198,119  
; PRIOR FILING DATE: 1998-11-23  
; PRIOR APPLICATION NUMBER: PCT/US99/01895  
; PRIOR FILING DATE: 1999-01-28  
; NUMBER OF SEQ ID NOS: 259  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11  
; LENGTH: 61  
; TYPE: PRT  
; ORGANISM: Tobacco  
US-09-996-140-11  
Query Match 3.9%; Score 7; DB 9; Length 61;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 95 LAYDKAA 101  
| | | | |  
Db 40 LAYDKAA 46  
RESULT 2  
US-09-996-140-134  
; Sequence 134, Application US/09996140  
; Patent No. US20020157136A1  
; GENERAL INFORMATION:  
; APPLICANT: Thomashow, Michael  
; APPLICANT: Stockinger, Eric  
; APPLICANT: Jaglo-Ottosen, Kirsten  
; APPLICANT: Gilmour, Sarah  
; APPLICANT: Zarka, Daniel  
; APPLICANT: Jiang, Cai-Zhong  
; TITLE OF INVENTION: PLANT HAVING ALTERED ENVIRONMENTAL  
; FILE REFERENCE: MB10040US  
; CURRENT APPLICATION NUMBER: US/09/996,140  
; CURRENT FILING DATE: 2001-11-26  
; PRIOR APPLICATION NUMBER: 09/706,270  
; PRIOR FILING DATE: 1996-09-04  
; PRIOR APPLICATION NUMBER: 09/018,233  
; PRIOR FILING DATE: 1998-02-03  
; PRIOR APPLICATION NUMBER: 09/017,816  
; PRIOR FILING DATE: 1998-02-03  
; PRIOR APPLICATION NUMBER: 09/018,227  
; PRIOR FILING DATE: 1998-02-03  
; PRIOR APPLICATION NUMBER: 09/018,234  
; PRIOR FILING DATE: 1998-02-03  
; PRIOR APPLICATION NUMBER: 09/198,119  
; PRIOR FILING DATE: 1998-11-23  
; PRIOR APPLICATION NUMBER: PCT/US99/01895  
; PRIOR FILING DATE: 1999-01-28  
; NUMBER OF SEQ ID NOS: 259  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 134  
; LENGTH: 63  
; TYPE: PRT  
; ORGANISM: Tobacco  
US-09-996-140-134  
Query Match 3.9%; Score 7; DB 9; Length 63;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 95 LAYDKAA 101  
| | | | |  
Db 42 LAYDKAA 48  
RESULT 3  
US-09-738-626-6630  
; Sequence 6630, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI

823	6	3.4	458	9	US-10-207-919-496	Sequence 496, App	896	6	3.4	620	9	US-10-140-470-324	Sequence 324, App
824	6	3.4	458	9	US-10-207-920-496	Sequence 496, App	897	6	3.4	620	9	US-09-906-838-73	Sequence 73, Appl
825	6	3.4	458	9	US-10-207-925-496	Sequence 496, App	898	6	3.4	620	9	US-09-907-613-73	Sequence 73, Appl
826	6	3.4	458	9	US-10-208-021-496	Sequence 496, App	899	6	3.4	620	9	US-09-907-942-73	Sequence 73, Appl
827	6	3.4	458	9	US-10-208-022-496	Sequence 496, App	900	6	3.4	620	9	US-10-175-748-324	Sequence 324, App
828	6	3.4	458	9	US-10-208-023-496	Sequence 496, App	901	6	3.4	620	9	US-10-176-918-324	Sequence 324, App
829	6	3.4	458	9	US-10-208-026-496	Sequence 496, App	902	6	3.4	620	9	US-10-176-921-324	Sequence 324, App
830	6	3.4	458	9	US-10-208-029-496	Sequence 496, App	903	6	3.4	620	9	US-10-137-865-324	Sequence 324, App
831	6	3.4	458	9	US-10-208-030-496	Sequence 496, App	904	6	3.4	620	9	US-10-140-474-324	Sequence 324, App
832	6	3.4	458	9	US-10-232-233-496	Sequence 496, App	905	6	3.4	620	9	US-09-904-820-73	Sequence 73, Appl
833	6	3.4	458	9	US-10-173-693-496	Sequence 496, App	906	6	3.4	620	9	US-09-904-859-73	Sequence 73, Appl
834	6	3.4	458	9	US-10-174-578-496	Sequence 496, App	907	6	3.4	620	9	US-09-909-204-73	Sequence 73, Appl
835	6	3.4	458	9	US-10-175-741-496	Sequence 496, App	908	6	3.4	620	9	US-10-142-431-324	Sequence 324, App
836	6	3.4	458	9	US-10-175-750-496	Sequence 496, App	909	6	3.4	620	9	US-10-143-114-324	Sequence 324, App
837	6	3.4	458	9	US-10-176-986-496	Sequence 496, App	910	6	3.4	620	9	US-09-904-786-73	Sequence 73, Appl
838	6	3.4	458	9	US-10-184-641-496	Sequence 496, App	911	6	3.4	620	9	US-09-906-646-73	Sequence 73, Appl
839	6	3.4	458	9	US-10-187-888-496	Sequence 496, App	912	6	3.4	620	9	US-09-906-700-73	Sequence 73, Appl
840	6	3.4	458	9	US-10-194-360-496	Sequence 496, App	913	6	3.4	620	9	US-10-140-002-324	Sequence 324, App
841	6	3.4	458	9	US-10-194-363-496	Sequence 496, App	914	6	3.4	620	9	US-09-902-903-73	Sequence 73, Appl
842	6	3.4	458	9	US-10-195-895-496	Sequence 496, App	915	6	3.4	620	9	US-09-903-749A-73	Sequence 73, Appl
843	6	3.4	458	9	US-10-195-898-496	Sequence 496, App	916	6	3.4	620	9	US-09-903-786-73	Sequence 73, Appl
844	6	3.4	458	9	US-10-196-759-496	Sequence 496, App	917	6	3.4	620	9	US-10-142-419-324	Sequence 324, App
845	6	3.4	458	9	US-10-199-302-496	Sequence 496, App	918	6	3.4	620	9	US-09-902-736-73	Sequence 73, Appl
846	6	3.4	458	9	US-10-201-323-496	Sequence 496, App	919	6	3.4	620	9	US-09-904-119-73	Sequence 73, Appl
847	6	3.4	458	9	US-10-205-510-496	Sequence 496, App	920	6	3.4	620	9	US-09-904-956-73	Sequence 73, Appl
848	6	3.4	458	9	US-10-205-891-496	Sequence 496, App	921	6	3.4	620	9	US-09-907-794-73	Sequence 73, Appl
849	6	3.4	458	9	US-10-205-904-496	Sequence 496, App	922	6	3.4	620	9	US-10-123-263-324	Sequence 324, App
850	6	3.4	458	9	US-10-206-917-496	Sequence 496, App	923	6	3.4	620	9	US-10-142-423-324	Sequence 324, App
851	6	3.4	458	9	US-10-207-923-496	Sequence 496, App	924	6	3.4	620	9	US-09-902-692-73	Sequence 73, Appl
852	6	3.4	458	9	US-10-207-924-496	Sequence 496, App	925	6	3.4	620	9	US-09-903-520-73	Sequence 73, Appl
853	6	3.4	458	9	US-10-208-028-496	Sequence 496, App	926	6	3.4	620	9	US-09-903-943-73	Sequence 73, Appl
854	6	3.4	458	9	US-10-121-062-496	Sequence 496, App	927	6	3.4	620	9	US-09-904-462-73	Sequence 73, Appl
855	6	3.4	458	9	US-10-175-753-496	Sequence 496, App	928	6	3.4	620	9	US-09-905-056-73	Sequence 73, Appl
856	6	3.4	458	9	US-10-180-553-496	Sequence 496, App	929	6	3.4	620	9	US-09-907-925-73	Sequence 73, Appl
857	6	3.4	458	9	US-10-201-327-496	Sequence 496, App	930	6	3.4	620	9	US-10-121-050-324	Sequence 324, App
858	6	3.4	458	9	US-10-173-696-496	Sequence 496, App	931	6	3.4	620	9	US-10-141-753-324	Sequence 324, App
859	6	3.4	458	9	US-10-183-003-496	Sequence 496, App	932	6	3.4	620	9	US-09-904-553-73	Sequence 73, Appl
860	6	3.4	458	9	US-10-183-016-496	Sequence 496, App	933	6	3.4	620	9	US-09-905-381-73	Sequence 73, Appl
861	6	3.4	458	12	US-10-052-586-496	Sequence 496, App	934	6	3.4	620	9	US-09-909-064-73	Sequence 324, App
862	6	3.4	478	12	US-10-007-693-69	Sequence 69, Appl	935	6	3.4	620	9	US-10-143-033-324	Sequence 324, App
863	6	3.4	478	9	US-10-105-717-2	Sequence 2, Appl	936	6	3.4	620	9	US-10-123-108-324	Sequence 324, App
864	6	3.4	481	10	US-09-158-180-2	Sequence 2, Appl	937	6	3.4	620	9	US-10-123-236-324	Sequence 324, App
865	6	3.4	495	10	US-09-823-356-4	Sequence 4, Appl	938	6	3.4	620	9	US-10-123-261-324	Sequence 324, App
866	6	3.4	496	10	US-09-815-242-13401	Sequence 13401, A	939	6	3.4	620	9	US-10-140-921-324	Sequence 324, App
867	6	3.4	502	10	US-09-815-242-5441	Sequence 5441, Ap	940	6	3.4	620	9	US-10-140-928-324	Sequence 324, App
868	6	3.4	502	10	US-09-815-242-12387	Sequence 12387, A	941	6	3.4	620	9	US-09-905-088-73	Sequence 73, Appl
869	6	3.4	502	10	US-09-815-242-12821	Sequence 12821, A	942	6	3.4	620	9	US-09-907-575-73	Sequence 73, Appl
870	6	3.4	504	9	US-09-972-268-8	Sequence 8, Appl	943	6	3.4	620	9	US-10-121-045-324	Sequence 324, App
871	6	3.4	510	9	US-09-959-845-4	Sequence 4, Appl	944	6	3.4	620	9	US-10-123-292-324	Sequence 324, App
872	6	3.4	510	9	US-09-972-268-10	Sequence 10, Appl	945	6	3.4	620	9	US-10-123-903-324	Sequence 324, App
873	6	3.4	510	9	US-09-972-268-12	Sequence 12, Appl	946	6	3.4	620	9	US-10-124-819-324	Sequence 324, App
874	6	3.4	510	9	US-09-972-268-18	Sequence 18, Appl	947	6	3.4	620	9	US-10-124-822-324	Sequence 324, App
875	6	3.4	512	9	US-09-983-204-14	Sequence 14, Appl	948	6	3.4	620	9	US-10-140-925-324	Sequence 324, App
876	6	3.4	542	9	US-09-972-268-2	Sequence 2, Appl	949	6	3.4	620	9	US-10-160-498-324	Sequence 324, App
877	6	3.4	549	9	US-09-959-845-2	Sequence 2, Appl	950	6	3.4	620	9	US-09-902-759-73	Sequence 73, Appl
878	6	3.4	549	9	US-09-972-268-4	Sequence 4, Appl	951	6	3.4	620	9	US-09-905-075-73	Sequence 73, Appl
879	6	3.4	549	9	US-09-972-268-6	Sequence 6, Appl	952	6	3.4	620	9	US-10-121-041-324	Sequence 324, App
880	6	3.4	549	9	US-09-972-268-17	Sequence 17, Appl	953	6	3.4	620	9	US-10-121-043-324	Sequence 324, App
881	6	3.4	569	9	US-09-906-419-39	Sequence 39, Appl	954	6	3.4	620	9	US-10-121-047-324	Sequence 324, App
882	6	3.4	594	9	US-10-255-969-10	Sequence 10, Appl	955	6	3.4	620	9	US-10-123-215-324	Sequence 324, App
883	6	3.4	594	12	US-10-139-282-10	Sequence 10, Appl	956	6	3.4	620	9	US-10-123-902-324	Sequence 324, App
884	6	3.4	595	9	US-09-972-268-14	Sequence 14, Appl	957	6	3.4	620	9	US-10-123-908-324	Sequence 324, App
885	6	3.4	614	10	US-09-803-589-4	Sequence 4, Appl	958	6	3.4	620	9	US-10-123-909-324	Sequence 324, App
886	6	3.4	620	9	US-09-905-291A-73	Sequence 73, Appl	959	6	3.4	620	9	US-10-123-910-324	Sequence 324, App
887	6	3.4	620	9	US-09-965-529-11	Sequence 11, Appl	960	6	3.4	620	9	US-10-124-813-324	Sequence 324, App
888	6	3.4	620	9	US-09-902-853-73	Sequence 73, Appl	961	6	3.4	620	9	US-10-124-817-324	Sequence 324, App
889	6	3.4	620	9	US-09-907-824-73	Sequence 73, Appl	962	6	3.4	620	9	US-10-124-824-324	Sequence 324, App
890	6	3.4	620	9	US-09-907-841-73	Sequence 73, Appl	963	6	3.4	620	9	US-10-125-922-324	Sequence 324, App
891	6	3.4	620	9	US-09-904-011-73	Sequence 73, Appl	964	6	3.4	620	9	US-10-125-924-324	Sequence 324, App
892	6	3.4	620	9	US-10-028-073-324	Sequence 324, App	965	6	3.4	620	9	US-10-127-825A-324	Sequence 324, App
893	6	3.4	620	9	US-09-906-742-73	Sequence 73, Appl	966	6	3.4	620	9	US-10-127-829A-324	Sequence 324, App
894	6	3.4	620	9	US-10-121-049-324	Sequence 324, App	967	6	3.4	620	9	US-10-127-835A-324	Sequence 324, App
895	6	3.4	620	9	US-10-123-904-324	Sequence 324, App	968	6	3.4	620	9	US-10-127-839A-324	Sequence 324, App





385	6	3.4	359	9	US-10-199-673-42	Sequence 42, Appl	458	6	3.4	359	9	US-10-180-553-42	Sequence 42, Appl
386	6	3.4	359	9	US-10-201-321-42	Sequence 42, Appl	459	6	3.4	359	9	US-10-201-327-42	Sequence 42, Appl
387	6	3.4	359	9	US-10-201-322-42	Sequence 42, Appl	460	6	3.4	359	9	US-09-978-757A-45	Sequence 45, Appl
388	6	3.4	359	9	US-10-201-326-42	Sequence 42, Appl	461	6	3.4	359	9	US-10-173-696-42	Sequence 42, Appl
389	6	3.4	359	9	US-10-201-532-42	Sequence 42, Appl	462	6	3.4	359	9	US-10-183-003-42	Sequence 42, Appl
390	6	3.4	359	9	US-10-201-533-42	Sequence 42, Appl	463	6	3.4	359	9	US-10-183-016-42	Sequence 42, Appl
391	6	3.4	359	9	US-10-201-535-42	Sequence 42, Appl	464	6	3.4	359	12	US-10-052-586-42	Sequence 42, Appl
392	6	3.4	359	9	US-10-201-769-42	Sequence 42, Appl	465	6	3.4	359	9	US-10-060-218A-6	Sequence 6, Appl
393	6	3.4	359	9	US-10-201-771-42	Sequence 42, Appl	466	6	3.4	374	9	US-09-847-208-68	Sequence 68, Appl
394	6	3.4	359	9	US-10-201-854-42	Sequence 42, Appl	467	6	3.4	387	9	US-09-972-268-16	Sequence 16, Appl
395	6	3.4	359	9	US-10-202-410-42	Sequence 42, Appl	468	6	3.4	393	9	US-09-977-418-12	Sequence 12, Appl
396	6	3.4	359	9	US-10-202-473-42	Sequence 42, Appl	469	6	3.4	393	9	US-09-977-033A-12	Sequence 12, Appl
397	6	3.4	359	9	US-10-202-474-42	Sequence 42, Appl	470	6	3.4	393	10	US-09-840-787-11	Sequence 11, Appl
398	6	3.4	359	9	US-10-205-503-42	Sequence 42, Appl	471	6	3.4	394	9	US-09-738-626-4096	Sequence 4096, Ap
399	6	3.4	359	9	US-10-205-512-42	Sequence 42, Appl	472	6	3.4	395	10	US-09-815-242-13892	Sequence 13892, A
400	6	3.4	359	9	US-10-205-892-42	Sequence 42, Appl	473	6	3.4	401	10	US-09-815-242-13828	Sequence 13828, A
401	6	3.4	359	9	US-10-205-894-42	Sequence 42, Appl	474	6	3.4	410	9	US-09-910-186A-32	Sequence 32, Appl
402	6	3.4	359	9	US-10-205-896-42	Sequence 42, Appl	475	6	3.4	411	9	US-09-866-050A-302	Sequence 302, App
403	6	3.4	359	9	US-10-205-898-42	Sequence 42, Appl	476	6	3.4	421	9	US-10-012-055-2	Sequence 2, Appl
404	6	3.4	359	9	US-10-205-901-42	Sequence 42, Appl	477	6	3.4	426	9	US-09-972-288-15	Sequence 15, Appl
405	6	3.4	359	9	US-10-205-903-42	Sequence 42, Appl	478	6	3.4	434	10	US-09-815-242-11827	Sequence 11827, A
406	6	3.4	359	9	US-10-206-909-42	Sequence 42, Appl	479	6	3.4	437	9	US-09-972-268-31	Sequence 31, Appl
407	6	3.4	359	9	US-10-206-910-42	Sequence 42, Appl	480	6	3.4	437	10	US-09-797-207-20	Sequence 20, Appl
408	6	3.4	359	9	US-10-206-911-42	Sequence 42, Appl	481	6	3.4	438	9	US-09-959-845-6	Sequence 6, Appl
409	6	3.4	359	9	US-10-206-912-42	Sequence 42, Appl	482	6	3.4	438	9	US-09-972-268-19	Sequence 19, Appl
410	6	3.4	359	9	US-10-206-913-42	Sequence 42, Appl	483	6	3.4	449	9	US-09-910-186A-14	Sequence 14, Appl
411	6	3.4	359	9	US-10-206-914-42	Sequence 42, Appl	484	6	3.4	456	10	US-09-997-664-92	Sequence 92, Appl
412	6	3.4	359	9	US-10-206-920-42	Sequence 42, Appl	485	6	3.4	458	9	US-10-174-590-496	Sequence 496, App
413	6	3.4	359	9	US-10-206-921-42	Sequence 42, Appl	486	6	3.4	458	9	US-10-176-758-496	Sequence 496, App
414	6	3.4	359	9	US-10-206-923-42	Sequence 42, Appl	487	6	3.4	458	9	US-10-175-737-496	Sequence 496, App
415	6	3.4	359	9	US-10-206-925-42	Sequence 42, Appl	488	6	3.4	458	9	US-10-173-706-496	Sequence 496, App
416	6	3.4	359	9	US-10-206-926-42	Sequence 42, Appl	489	6	3.4	458	9	US-10-175-728-496	Sequence 496, App
417	6	3.4	359	9	US-10-206-927-42	Sequence 42, Appl	490	6	3.4	458	9	US-10-175-752-496	Sequence 496, App
418	6	3.4	359	9	US-10-207-916-42	Sequence 42, Appl	491	6	3.4	458	9	US-10-176-482-496	Sequence 496, App
419	6	3.4	359	9	US-10-207-917-42	Sequence 42, Appl	492	6	3.4	458	9	US-10-176-757-496	Sequence 496, App
420	6	3.4	359	9	US-10-207-918-42	Sequence 42, Appl	493	6	3.4	458	9	US-10-176-913-496	Sequence 496, App
421	6	3.4	359	9	US-10-207-919-42	Sequence 42, Appl	494	6	3.4	458	9	US-10-180-552-496	Sequence 496, App
422	6	3.4	359	9	US-10-207-920-42	Sequence 42, Appl	495	6	3.4	458	9	US-10-180-557-496	Sequence 496, App
423	6	3.4	359	9	US-10-207-925-42	Sequence 42, Appl	496	6	3.4	458	9	US-10-173-700-496	Sequence 496, App
424	6	3.4	359	9	US-10-208-021-42	Sequence 42, Appl	497	6	3.4	458	9	US-10-174-572-496	Sequence 496, App
425	6	3.4	359	9	US-10-208-022-42	Sequence 42, Appl	498	6	3.4	458	9	US-10-174-579-496	Sequence 496, App
426	6	3.4	359	9	US-10-208-023-42	Sequence 42, Appl	499	6	3.4	458	9	US-10-174-582-496	Sequence 496, App
427	6	3.4	359	9	US-10-208-026-42	Sequence 42, Appl	500	6	3.4	458	9	US-10-174-588-496	Sequence 496, App
428	6	3.4	359	9	US-10-208-029-42	Sequence 42, Appl	501	6	3.4	458	9	US-10-175-739-496	Sequence 496, App
429	6	3.4	359	9	US-10-208-030-42	Sequence 42, Appl	502	6	3.4	458	9	US-10-175-740-496	Sequence 496, App
430	6	3.4	359	9	US-10-232-232-42	Sequence 42, Appl	503	6	3.4	458	9	US-10-173-693-496	Sequence 496, App
431	6	3.4	359	9	US-09-978-193A-45	Sequence 45, Appl	504	6	3.4	458	9	US-10-176-488-496	Sequence 496, App
432	6	3.4	359	9	US-10-013-929A-45	Sequence 45, Appl	505	6	3.4	458	9	US-10-176-492-496	Sequence 496, App
433	6	3.4	359	9	US-10-016-177A-45	Sequence 45, Appl	506	6	3.4	458	9	US-10-176-747-496	Sequence 496, App
434	6	3.4	359	9	US-10-173-693-42	Sequence 42, Appl	507	6	3.4	458	9	US-10-176-750-496	Sequence 496, App
435	6	3.4	359	9	US-10-174-578-42	Sequence 42, Appl	508	6	3.4	458	9	US-10-176-985-496	Sequence 496, App
436	6	3.4	359	9	US-10-175-741-42	Sequence 42, Appl	509	6	3.4	458	9	US-10-176-987-496	Sequence 496, App
437	6	3.4	359	9	US-10-175-750-42	Sequence 42, Appl	510	6	3.4	458	9	US-10-176-991-496	Sequence 496, App
438	6	3.4	359	9	US-10-176-986-42	Sequence 42, Appl	511	6	3.4	458	9	US-10-176-992-496	Sequence 496, App
439	6	3.4	359	9	US-10-184-641-42	Sequence 42, Appl	512	6	3.4	458	9	US-10-176-993-496	Sequence 496, App
440	6	3.4	359	9	US-10-187-888-42	Sequence 42, Appl	513	6	3.4	458	9	US-10-184-658-496	Sequence 496, App
441	6	3.4	359	9	US-10-194-360-42	Sequence 42, Appl	514	6	3.4	458	9	US-10-173-695-496	Sequence 496, App
442	6	3.4	359	9	US-10-194-365-42	Sequence 42, Appl	515	6	3.4	458	9	US-10-173-697-496	Sequence 496, App
443	6	3.4	359	9	US-10-195-895-42	Sequence 42, Appl	516	6	3.4	458	9	US-10-173-705-496	Sequence 496, App
444	6	3.4	359	9	US-10-195-898-42	Sequence 42, Appl	517	6	3.4	458	9	US-10-174-576-496	Sequence 496, App
445	6	3.4	359	9	US-10-196-759-42	Sequence 42, Appl	518	6	3.4	458	9	US-10-174-585-496	Sequence 496, App
446	6	3.4	359	9	US-10-199-302-42	Sequence 42, Appl	519	6	3.4	458	9	US-10-174-586-496	Sequence 496, App
447	6	3.4	359	9	US-10-201-323-42	Sequence 42, Appl	520	6	3.4	458	9	US-10-175-747-496	Sequence 496, App
448	6	3.4	359	9	US-10-203-510-42	Sequence 42, Appl	521	6	3.4	458	9	US-10-176-481-496	Sequence 496, App
449	6	3.4	359	9	US-10-205-891-42	Sequence 42, Appl	522	6	3.4	458	9	US-10-176-485-496	Sequence 496, App
450	6	3.4	359	9	US-10-205-904-42	Sequence 42, Appl	523	6	3.4	458	9	US-10-176-487-496	Sequence 496, App
451	6	3.4	359	9	US-10-206-917-42	Sequence 42, Appl	524	6	3.4	458	9	US-10-176-493-496	Sequence 496, App
452	6	3.4	359	9	US-10-207-923-42	Sequence 42, Appl	525	6	3.4	458	9	US-10-176-756-496	Sequence 496, App
453	6	3.4	359	9	US-10-207-924-42	Sequence 42, Appl	526	6	3.4	458	9	US-10-176-911-496	Sequence 496, App
454	6	3.4	359	9	US-10-208-028-42	Sequence 42, Appl	527	6	3.4	458	9	US-10-176-919-496	Sequence 496, App
455	6	3.4	359	9	US-09-999-830A-45	Sequence 45, Appl	528	6	3.4	458	9	US-10-176-925-496	Sequence 496, App
456	6	3.4	359	9	US-10-121-062-42	Sequence 42, Appl	529	6	3.4	458	9	US-10-176-978-496	Sequence 496, App
457	6	3.4	359	9	US-10-175-753-42	Sequence 42, Appl	530	6	3.4	458	9	US-10-175-510-496	Sequence 496, App







GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: May 12, 2003, 09:48:56 ; Search time 20 seconds  
(without alignments)  
823.630 Million cell updates/sec

Title: US-09-995-493-52

Perfect score: 179

Sequence: 1 QHNGVLGPYICKSLTKLP.....ADYTTKATNLYGLNLNRYE 179

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 349150 seqs, 92025710 residues

Word size : 0

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published\_Applications\_AA.\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	7	3.9	61	9	US-09-996-140-11
2	7	3.9	63	9	US-09-996-140-134
3	7	3.9	93	9	US-09-738-626-6630
4	7	3.9	218	9	US-10-278-173-150
5	7	3.9	334	9	US-10-278-173-82
6	7	3.9	511	10	US-09-881-752A-110
7	7	3.9	795	9	US-09-738-626-3810
8	6	3.4	8	9	US-10-024-918-10
9	6	3.4	21	9	US-10-150-111-181
10	6	3.4	24	9	US-10-097-065-537
11	6	3.4	34	10	US-09-864-761-36829
12	6	3.4	37	10	US-09-864-761-40107
13	6	3.4	59	9	US-10-122-822-25
14	6	3.4	60	10	US-09-867-550-1856
15	6	3.4	64	10	US-09-864-761-34896
16	6	3.4	64	10	US-09-867-550-904
17	6	3.4	68	10	US-09-864-761-41976
18	6	3.4	68	10	US-09-864-761-46121
19	6	3.4	69	10	US-09-925-300-1548

20	6	3.4	76	9	US-10-079-854-110	Sequence 110, App
21	6	3.4	76	10	US-09-764-878-110	Sequence 110, App
22	6	3.4	80	10	US-09-893-737-12	Sequence 12, Appl
23	6	3.4	82	9	US-09-738-626-6220	Sequence 6220, Ap
24	6	3.4	90	9	US-09-866-050A-134	Sequence 134, App
25	6	3.4	90	10	US-09-864-761-43578	Sequence 43578, A
26	6	3.4	91	10	US-09-864-761-43033	Sequence 43033, A
27	6	3.4	93	10	US-09-867-550-612	Sequence 612, App
28	6	3.4	102	9	US-10-300-616-49	Sequence 49, Appl
29	6	3.4	104	10	US-09-864-761-41522	Sequence 41522, A
30	6	3.4	110	9	US-10-092-154-844	Sequence 844, App
31	6	3.4	110	10	US-09-764-847-844	Sequence 844, App
32	6	3.4	111	10	US-10-125-258-8	Sequence 8, Appl
33	6	3.4	132	9	US-09-738-626-4318	Sequence 4318, Ap
34	6	3.4	137	9	US-09-862-540-51	Sequence 51, Appl
35	6	3.4	142	9	US-10-102-808-799	Sequence 799, App
36	6	3.4	148	9	US-10-125-258-10	Sequence 10, Appl
37	6	3.4	149	9	US-09-983-802-196	Sequence 196, App
38	6	3.4	153	10	US-09-925-300-1688	Sequence 1688, Ap
39	6	3.4	162	1	US-08-781-986A-5220	Sequence 5220, Ap
40	6	3.4	168	9	US-09-981-876-184	Sequence 184, App
41	6	3.4	168	9	US-09-148-545-184	Sequence 184, App
42	6	3.4	175	9	US-10-150-111-173	Sequence 173, App
43	6	3.4	183	9	US-09-895-298-70	Sequence 70, Appl
44	6	3.4	191	9	US-10-092-154-831	Sequence 831, App
45	6	3.4	191	10	US-09-764-847-831	Sequence 831, Appl
46	6	3.4	196	9	US-10-125-258-4	Sequence 4, Appl
47	6	3.4	217	10	US-09-764-853-664	Sequence 664, App
48	6	3.4	223	9	US-10-097-065-325	Sequence 325, App
49	6	3.4	231	9	US-10-122-822-1	Sequence 1, Appl
50	6	3.4	233	9	US-10-125-258-18	Sequence 18, Appl
51	6	3.4	253	9	US-10-150-111-84	Sequence 84, Appl
52	6	3.4	262	9	US-10-235-674-14	Sequence 14, Appl
53	6	3.4	262	10	US-09-263-689-14	Sequence 14, Appl
54	6	3.4	276	10	US-09-765-272-134	Sequence 134, App
55	6	3.4	278	10	US-09-925-300-1323	Sequence 1323, Ap
56	6	3.4	282	9	US-10-150-111-172	Sequence 172, App
57	6	3.4	305	9	US-10-102-806-624	Sequence 624, App
58	6	3.4	325	9	US-09-738-626-6035	Sequence 6035, Ap
59	6	3.4	325	9	US-09-746-660A-4	Sequence 4, Appl
60	6	3.4	326	9	US-09-957-187-75	Sequence 75, Appl
61	6	3.4	329	9	US-10-159-901-33	Sequence 33, Appl
62	6	3.4	329	9	US-10-159-901-34	Sequence 34, Appl
63	6	3.4	329	9	US-10-159-901-55	Sequence 55, Appl
64	6	3.4	340	10	US-09-250-883-17	Sequence 17, Appl
65	6	3.4	359	9	US-09-978-295A-45	Sequence 45, Appl
66	6	3.4	359	9	US-09-978-697-45	Sequence 45, Appl
67	6	3.4	359	9	US-09-978-192A-45	Sequence 45, Appl
68	6	3.4	359	9	US-09-999-832A-45	Sequence 45, Appl
69	6	3.4	359	9	US-09-978-189-45	Sequence 45, Appl
70	6	3.4	359	9	US-10-174-590-42	Sequence 42, Appl
71	6	3.4	359	9	US-10-176-758-42	Sequence 42, Appl
72	6	3.4	359	9	US-10-175-737-42	Sequence 42, Appl
73	6	3.4	359	9	US-10-173-706-42	Sequence 42, Appl
74	6	3.4	359	9	US-10-175-738-42	Sequence 42, Appl
75	6	3.4	359	9	US-10-175-753-42	Sequence 42, Appl
76	6	3.4	359	9	US-10-176-482-42	Sequence 42, Appl
77	6	3.4	359	9	US-10-176-482-42	Sequence 42, Appl
78	6	3.4	359	9	US-10-176-757-42	Sequence 42, Appl
79	6	3.4	359	9	US-10-176-913-42	Sequence 42, Appl
80	6	3.4	359	9	US-10-180-552-42	Sequence 42, Appl
81	6	3.4	359	9	US-10-180-557-42	Sequence 42, Appl
82	6	3.4	359	9	US-10-173-700-42	Sequence 42, Appl
83	6	3.4	359	9	US-10-174-572-42	Sequence 42, Appl
84	6	3.4	359	9	US-10-174-579-42	Sequence 42, Appl
85	6	3.4	359	9	US-10-174-582-42	Sequence 42, Appl
86	6	3.4	359	9	US-10-174-588-42	Sequence 42, Appl
87	6	3.4	359	9	US-10-175-739-42	Sequence 42, Appl
88	6	3.4	359	9	US-10-175-740-42	Sequence 42, Appl
89	6	3.4	359	9	US-10-175-743-42	Sequence 42, Appl
90	6	3.4	359	9	US-10-176-488-42	Sequence 42, Appl
91	6	3.4	359	9	US-10-176-492-42	Sequence 42, Appl
92	6	3.4	359	9	US-10-176-747-42	Sequence 42, Appl
					US-10-176-750-42	Sequence 42, Appl

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 09:46:15 ; Search time 17 Seconds  
(without alignments)  
1012.239 Million cell updates/sec

Title: US-09-995-493-52

Perfect score: 179

Sequence: 1 QHNGVLGPYIGKSLTKLP.....ADYTKRATNLYGLNLNRYF 179

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13	7.3	450	2 B40183	outer membrane pro
2	13	7.3	455	2 A40183	outer membrane pro
3	13	7.3	459	2 A28787	outer membrane pro
4	13	7.3	459	2 F64065	outer membrane pro
5	9	5.0	412	2 G82406	long-chain fatty a
6	8	4.5	72	2 H69448	heme d1 biosynthes
7	8	4.5	417	2 E90844	probable factor li
8	8	4.5	417	2 D85702	hypothetical prote
9	8	4.5	417	2 A64869	ychp protein - Esc
10	8	4.5	633	2 F75107	abc transporter, A
11	7	3.9	66	2 D98068	degenerate transpo
12	7	3.9	146	1 MYUY	myoglobin [validat
13	7	3.9	168	2 S00942	hypothetical prote
14	7	3.9	180	2 AB2849	hypothetical prote
15	7	3.9	218	2 F84748	probable AP2 domai
16	7	3.9	233	2 T02590	DNA binding protei
17	7	3.9	260	1 CRMS2	carbonate dehydrat
18	7	3.9	261	2 E86354	hypothetical prote
19	7	3.9	274	2 B71877	hypothetical prote
20	7	3.9	308	2 F64498	selenium donor pro
21	7	3.9	336	2 T35643	glycerol-3-phospha
22	7	3.9	356	2 B30338	late competence pr
23	7	3.9	363	2 C84179	aminopeptidase hom
24	7	3.9	413	2 A70795	hypothetical prote
25	7	3.9	423	2 AH0334	probable long-chain
26	7	3.9	437	2 AE0805	long-chain fatty a
27	7	3.9	448	2 F65007	long-chain fatty a
28	7	3.9	448	2 C91032	long-chain fatty a
29	7	3.9	448	2 D85876	long-chain fatty a

30	7	3.9	511	2 A71859	probable outer mem
31	7	3.9	519	2 F64564	hypothetical prote
32	7	3.9	519	2 S77572	oligopeptide trans
33	7	3.9	526	2 T08541	hypothetical prote
34	7	3.9	528	2 AI0359	iron(III)-transpor
35	7	3.9	547	2 B45808	B-lymphocyte antig
36	7	3.9	556	2 S39297	fiber protein - hu
37	7	3.9	556	2 A44441	B-cell antigen CD1
38	7	3.9	671	2 AD1666	DNA ligase homolog
39	7	3.9	693	2 T47474	hypothetical prote
40	7	3.9	734	2 T02196	hypothetical prote
41	7	3.9	849	2 S61631	probable membrane
42	7	3.9	859	2 B64430	DNA-directed RNA p
43	7	3.9	865	1 C64737	outer membrane ush
44	7	3.9	977	2 A30111	alpha-adaptin A -
45	7	3.9	1094	2 C59434	KIAA1688 protein l
46	7	3.9	1641	2 D82704	conserved hypothet
47	7	3.9	2340	2 I48310	kinase-related pro
48	7	3.9	3216	2 C90538	hypothetical prote
49	6	3.4	49	2 S49365	Ig mu chain transm
50	6	3.4	52	2 I37971	zinc finger protei
51	6	3.4	65	2 D90082	hypothetical prote
52	6	3.4	65	2 E90101	hypothetical prote
53	6	3.4	65	2 D90118	hypothetical prote
54	6	3.4	65	2 H90137	hypothetical prote
55	6	3.4	65	2 A99117	hypothetical prote
56	6	3.4	72	2 H82334	slxX protein VC035
57	6	3.4	77	2 C86842	30S ribosomal prot
58	6	3.4	81	2 E90324	hypothetical prote
59	6	3.4	82	2 S52460	hypothetical prote
60	6	3.4	84	2 E82709	hypothetical prote
61	6	3.4	86	2 T10047	ribosomal protein
62	6	3.4	86	2 G70684	probable ribosomal
63	6	3.4	86	2 C86269	protein F21F23.5 [
64	6	3.4	93	2 G69361	transposase IS240-
65	6	3.4	97	2 S72333	hypothetical prote
66	6	3.4	104	1 S73130	ribosomal protein
67	6	3.4	105	2 T10117	flmbriata-associat
68	6	3.4	108	2 C72457	hypothetical prote
69	6	3.4	112	2 A60111	sporozoite antigen
70	6	3.4	119	2 AF0229	phage-related prot
71	6	3.4	120	2 T46735	transcription regu
72	6	3.4	122	2 A53878	type E neurotoxin
73	6	3.4	123	1 R5RT35	ribosomal protein
74	6	3.4	123	2 G01477	ribosomal protein
75	6	3.4	124	2 A71122	hypothetical prote
76	6	3.4	124	2 F72516	hypothetical prote
77	6	3.4	124	2 E95961	conserved hypothet
78	6	3.4	129	2 E90331	hypothetical prote
79	6	3.4	133	2 C90662	hypothetical prote
80	6	3.4	133	2 A85513	hypothetical prote
81	6	3.4	134	2 A65133	hypothetical 15.2
82	6	3.4	139	2 P80396	hypothetical prote
83	6	3.4	140	2 C83199	hypothetical prote
84	6	3.4	142	2 T30401	hypothetical prote
85	6	3.4	143	2 G84168	hypothetical prote
86	6	3.4	146	2 F72505	hypothetical prote
87	6	3.4	149	2 T02709	probable kinetecho
88	6	3.4	151	2 T17529	SAP1 protein homol
89	6	3.4	153	2 G70077	hypothetical prote
90	6	3.4	154	2 C82261	conserved hypothet
91	6	3.4	156	2 AG0716	probable heat shoc
92	6	3.4	158	2 D72691	hypothetical prote
93	6	3.4	161	2 H84458	hypothetical prote
94	6	3.4	162	2 T12538	hypothetical prote
95	6	3.4	173	2 F97105	HD family hydrolas
96	6	3.4	173	2 I51269	lutropin receptor
97	6	3.4	177	2 B83906	hypothetical prote
98	6	3.4	179	2 S54222	outer surface prot
99	6	3.4	180	2 S54189	outer surface prot
100	6	3.4	181	2 D72060	conserved hypothet
101	6	3.4	181	2 A86564	CT469 hypothetical
102	6	3.4	182	2 T15754	hypothetical prote

103	6	3.4	184	2	AB1731	hypothetical prote	176	6	3.4	259	2	B85534	rRNA methylase [Im
104	6	3.4	185	2	D83658	stage v sporulatio	177	6	3.4	259	2	G85363	hypothetical prote
105	6	3.4	186	2	S14840	dihydrofolate redu	178	6	3.4	260	2	A95927	probable transcrip
106	6	3.4	187	1	AGEC2	p-aminobenzoate sy	179	6	3.4	261	2	T35708	hydrolase - Strept
107	6	3.4	187	2	S09636	pABA protein - Sal	180	6	3.4	261	2	T72426	iron-sulfur cluste
108	6	3.4	187	2	C91155	p-aminobenzoate sy	181	6	3.4	262	2	D71068	probable pseudouri
109	6	3.4	187	2	H86000	p-aminobenzoate sy	182	6	3.4	262	2	A54889	IgE-binding protei
110	6	3.4	187	2	AD1002	para-aminobenzoate	183	6	3.4	263	2	C64944	transcription regu
111	6	3.4	188	2	D75499	probable acetyltra	184	6	3.4	263	2	A89946	transcription regu
112	6	3.4	189	2	S02669	outer surface prot	185	6	3.4	263	2	AB0209	probable regulator
113	6	3.4	189	2	E90408	conserved hypothet	186	6	3.4	263	2	E85794	Ic1R-family transc
114	6	3.4	189	2	B72348	hypothetical prote	187	6	3.4	263	2	AB0728	probable regulator
115	6	3.4	190	2	H75478	conserved hypothet	188	6	3.4	263	2	T32081	transcription regu
116	6	3.4	190	2	G69339	conserved hypothet	189	6	3.4	264	2	A28651	hypothetical prote
117	6	3.4	191	2	G71710	sco2 protein precu	190	6	3.4	264	2	A45983	lactose-specific
118	6	3.4	191	2	S02771	outer surface prot	191	6	3.4	264	2	AG3466	transcription regu
119	6	3.4	191	2	S02778	outer surface prot	192	6	3.4	266	2	C70788	hypothetical prote
120	6	3.4	192	2	H69387	conserved hypothet	193	6	3.4	267	1	S33964	orotidine-5'-phosp
121	6	3.4	192	2	AD1036	probable phage bas	194	6	3.4	267	2	T37689	hypothetical prote
122	6	3.4	192	2	AF0928	probable phage bas	195	6	3.4	267	2	D95089	hypothetical prote
123	6	3.4	194	2	T06831	endopeptidase Clp	196	6	3.4	267	2	G97956	peptidyl-prolyl cl
124	6	3.4	194	2	S02777	outer surface prot	197	6	3.4	269	2	F95106	Cof family protein
125	6	3.4	195	2	T35651	probable methylase	198	6	3.4	269	2	H97974	conserved hypothet
126	6	3.4	196	2	G83757	ribonuclease H-rel	199	6	3.4	270	2	G84226	hypothetical prote
127	6	3.4	199	2	B97705	sco2 protein precu	200	6	3.4	270	2	T37809	probable zinc flng
128	6	3.4	203	2	S54224	outer surface prot	201	6	3.4	272	2	S33622	ADP6 protein - soy
129	6	3.4	204	2	B87704	heme exporter prot	202	6	3.4	277	2	C84273	archaeal nucleolar
130	6	3.4	204	2	B72370	hypothetical prote	203	6	3.4	279	2	A99517	ABC transporter at
131	6	3.4	206	2	E81221	probable integral	204	6	3.4	279	2	F97024	hypothetical prote
132	6	3.4	206	2	S76502	hypothetical prote	205	6	3.4	281	2	T35932	probable inositol
133	6	3.4	207	2	S71452	trypsin/chymotryps	206	6	3.4	281	2	T04858	hypothetical prote
134	6	3.4	207	2	S69924	outer surface prot	207	6	3.4	283	2	H64764	probable taurine d
135	6	3.4	208	2	C90551	hypothetical prote	208	6	3.4	283	2	F90681	taurine dioxigenas
136	6	3.4	208	2	H91033	probable resolvase	209	6	3.4	283	2	B85532	high-affinity tran
137	6	3.4	208	2	A85878	probable resolvase	210	6	3.4	283	2	F98243	ABC transporter, m
138	6	3.4	209	2	I40273	outer surface prot	211	6	3.4	283	2	AF3042	hypothetical prote
139	6	3.4	209	2	I40270	outer surface prot	212	6	3.4	285	2	G72414	spore germination
140	6	3.4	209	2	T34027	hypothetical prote	213	6	3.4	288	1	B48583	hypothetical prote
141	6	3.4	211	2	I40277	outer surface prot	214	6	3.4	288	2	T32128	hypothetical prote
142	6	3.4	211	2	S89928	outer surface prot	215	6	3.4	289	2	S35022	probable N-acetyl
143	6	3.4	212	2	T06015	hypothetical prote	216	6	3.4	290	2	S66072	yabG protein - Bac
144	6	3.4	213	1	RD8EHS	dihydrofolate redu	217	6	3.4	290	2	G36789	hypothetical prote
145	6	3.4	216	1	S74616	uracil phosphoribo	218	6	3.4	291	1	F36516	3-methylcatechol 2
146	6	3.4	216	2	S05575	sporozoite antigen	219	6	3.4	291	2	A12468	hypothetical prote
147	6	3.4	217	2	S61625	hypothetical prote	220	6	3.4	294	2	D90578	elongation factor
148	6	3.4	224	2	T03969	uracil phosphoribo	221	6	3.4	295	1	D72021	probable 3',5'-cyc
149	6	3.4	228	2	D64672	ABC transporter, A	222	6	3.4	295	1	G86602	probable 3',5'-cyc
150	6	3.4	228	2	G71842	probable ABC trans	223	6	3.4	296	2	T45930	uracil phosphoribo
151	6	3.4	228	2	G64573	protein phosphatas	224	6	3.4	297	1	MNVNAV	nonstructural prot
152	6	3.4	229	2	A82959	hypothetical prote	225	6	3.4	297	2	S21085	salivary glue prot
153	6	3.4	233	1	EWK9BC	atactin precursor,	226	6	3.4	297	2	S13706	nonstructural phos
154	6	3.4	234	2	T07686	transcription fact	227	6	3.4	299	2	A95878	probable N-acetyl
155	6	3.4	234	2	A75587	probable urea/short	228	6	3.4	299	2	AG2340	hypothetical prote
156	6	3.4	236	2	T02432	ethylene-responsiv	229	6	3.4	301	2	D82040	cysG protein VC272
157	6	3.4	236	2	AF2288	hypothetical prote	230	6	3.4	302	2	B49889	regulatory protein
158	6	3.4	239	2	T36870	probable DNA hydro	231	6	3.4	303	2	T04541	hypothetical prote
159	6	3.4	242	2	JC4300	galectin-3 - rabbi	232	6	3.4	304	2	T34845	hypothetical prote
160	6	3.4	243	2	S06098	hypothetical 26.6k	233	6	3.4	305	1	H75285	probable phosphos
161	6	3.4	243	2	H82253	ribonuclease T VC1	234	6	3.4	305	2	T03155	ribonucleoside-dip
162	6	3.4	244	2	AB3524	transcription regu	235	6	3.4	305	2	T08836	probable metallopro
163	6	3.4	245	2	A54909	carbohydrate-blndi	236	6	3.4	305	2	A10476	oxidative stress t
164	6	3.4	245	2	G75422	hypothetical prote	237	6	3.4	306	2	S17713	transcription regu
165	6	3.4	247	1	S18532	hypothetical prote	238	6	3.4	308	2	A84566	Ac-like transposas
166	6	3.4	247	1	G84280	probable phosphos	239	6	3.4	308	2	T46023	hypothetical prote
167	6	3.4	247	1	A55910	H+-transporting tw	240	6	3.4	309	2	T21195	hypothetical prote
168	6	3.4	250	2	B71859	probable oxidoredu	241	6	3.4	310	2	T36542	hypothetical prote
169	6	3.4	250	2	E64564	short chain alchoho	242	6	3.4	313	1	H69510	conserved hypothet
170	6	3.4	253	2	A83034	probable short-cha	243	6	3.4	313	2	JC5475	pollen allergen 9
171	6	3.4	253	2	H70217	conserved hypothet	244	6	3.4	314	2	AF3268	signal recognition
172	6	3.4	254	2	AG3331	cobalt transport p	245	6	3.4	314	2	T23301	hypothetical prote
173	6	3.4	257	2	T23332	hypothetical prote	246	6	3.4	317	2	JC5696	prolyl aminopeptid
174	6	3.4	258	2	S75950	hypothetical prote	247	6	3.4	319	2	A70959	probable aspartate
175	6	3.4	259	2	B72089	ABC transporter, A	248	6	3.4	320	2	T36062	probable ABC-type

249	6	3.4	321	2	D86975	probable aspartate	322	6	3.4	401	2	E84367	alcohol dehydrogen
250	6	3.4	321	2	S54403	1-phosphatidylinos	323	6	3.4	404	2	H90983	probable glycosyl
251	6	3.4	323	2	T05478	peroxidase (EC 1.1	324	6	3.4	404	2	C85829	glycosyl transfera
252	6	3.4	324	2	F69504	proliferating-cell	325	6	3.4	405	2	A55967	1-phosphatidylinos
253	6	3.4	325	2	T05356	hypothetical prote	326	6	3.4	406	2	S7217	1-phosphatidylinos
254	6	3.4	326	2	A41862	C-S lyase (ORF326)	327	6	3.4	406	2	E81300	probable glucose-6
255	6	3.4	327	2	A64613	conserved hypotnet	328	6	3.4	407	2	B71690	hypothetical prote
256	6	3.4	327	2	B71900	hypothetical prote	329	6	3.4	411	2	B71500	probable amino aci
257	6	3.4	329	1	JC4251	D-xylose 1-dehydro	330	6	3.4	413	2	A97542	D-ala-D-ala-carbox
258	6	3.4	331	2	A99179	conserved hypotnet	331	6	3.4	414	2	D86484	hypothetical prote
259	6	3.4	333	2	G87426	ADP-L-glycero-D-ma	332	6	3.4	415	2	B86434	protein T17H7.13 [
260	6	3.4	333	2	T40923	pyridoxal reductas	333	6	3.4	418	2	H83128	probable porin PA4
261	6	3.4	334	2	T19127	hypothetical prote	334	6	3.4	420	2	C69128	glutamate-1-semial
262	6	3.4	335	2	D38532	hyPB protein - Rho	335	6	3.4	420	2	I49708	GTP-binding protei
263	6	3.4	336	2	T49112	aspartic proteinas	336	6	3.4	422	2	A12563	hypothetical prote
264	6	3.4	337	2	C84297	hypothetical prote	337	6	3.4	423	2	I59463	keratin, type I, c
265	6	3.4	337	2	AH3207	transcription regu	338	6	3.4	424	2	B81936	tetrahydrofolylpol
266	6	3.4	345	2	B97066	aldose-1-epimerase	339	6	3.4	424	2	C81170	folypolyglutamate
267	6	3.4	345	2	T48313	hypothetical prote	340	6	3.4	424	2	S63583	tetrahydrofolylpol
268	6	3.4	347	2	B85432	hypothetical prote	341	6	3.4	424	2	T43498	hypothetical prote
269	6	3.4	349	2	D70141	oligopeptide ABC t	342	6	3.4	425	2	AG2787	long-chain fatty a
270	6	3.4	350	2	C96532	peroxidase Atp5a 1	343	6	3.4	428	2	E87573	chlorohydrolase [1
271	6	3.4	352	2	A49210	heat shock-protein	344	6	3.4	429	2	A97567	hypothetical prote
272	6	3.4	352	2	C82294	oxidoreductase fas	345	6	3.4	430	2	T34627	probable electron
273	6	3.4	352	2	T18794	hypothetical prote	346	6	3.4	430	2	E83870	hypothetical prote
274	6	3.4	356	2	F82069	conserved hypotnet	347	6	3.4	431	2	T24101	hypothetical prote
275	6	3.4	356	2	D82643	conserved hypotnet	348	6	3.4	431	2	E82766	conserved hypotnet
276	6	3.4	357	2	F82878	XAA-PRO aminopepti	349	6	3.4	432	2	B82250	long-chain fatty a
277	6	3.4	359	2	A75438	recF protein - Del	350	6	3.4	432	2	A85436	APETALA2 protein l
278	6	3.4	360	2	I51663	arginase 1 - Afric	351	6	3.4	444	2	PD0001	outer membrane pro
279	6	3.4	360	2	I51664	arginase 2 - Afric	352	6	3.4	445	2	T24105	ABC transporter UU
280	6	3.4	360	2	I51665	arginase 3 - Afric	353	6	3.4	446	2	D82879	probable guanine d
281	6	3.4	362	2	C64807	ybgO protein - Esc	354	6	3.4	434	2	D83456	ORF MSV181 hypotne
282	6	3.4	362	2	T29552	hypothetical prote	355	6	3.4	442	2	T28342	conserved hypotnet
283	6	3.4	363	2	G71544	hypothetical prote	356	6	3.4	444	2	D95889	protein-glutamine
284	6	3.4	364	2	D70164	heat shock protein	357	6	3.4	445	2	T24105	hypothetical prote
285	6	3.4	365	2	E83710	hypothetical prote	358	6	3.4	446	2	T03981	APETALA2-like prot
286	6	3.4	366	2	T48501	hypothetical prote	359	6	3.4	447	2	T02547	hypothetical prote
287	6	3.4	368	1	I64231	high affinity tran	360	6	3.4	447	2	C75413	probable proteinas
288	6	3.4	369	2	E75620	hypothetical prote	361	6	3.4	448	2	F84780	hypothetical prote
289	6	3.4	370	2	E82369	thiH protein VC006	362	6	3.4	448	2	F95122	protein kinase, pr
290	6	3.4	371	2	A71378	probable DNA polym	363	6	3.4	448	2	G70977	hypothetical prote
291	6	3.4	371	2	F70634	hypothetical prote	364	6	3.4	450	2	E70681	probable trpE - My
292	6	3.4	371	2	AH2172	hypothetical prote	365	6	3.4	453	2	G96573	protein F12M16.27-
293	6	3.4	373	2	D64106	dprA protein - Hae	366	6	3.4	453	2	S53995	ToxX protein - Pse
294	6	3.4	374	2	JC2124	major allergen Cry	367	6	3.4	454	2	T24455	hypothetical prote
295	6	3.4	374	2	JC2123	major allergen Cry	368	6	3.4	454	2	G97125	probable inactivat
296	6	3.4	377	2	T47439	hypothetical prote	369	6	3.4	455	2	S71344	purinergic recepto
297	6	3.4	378	2	E70786	hypothetical prote	370	6	3.4	460	2	B34283	H+-transporting tw
298	6	3.4	383	2	D86781	L-lactate oxidase	371	6	3.4	460	2	T45108	H+-transporting tw
299	6	3.4	383	2	S56906	GCPI4 protein - ye	372	6	3.4	462	2	C83747	dihydrolipoamide d
300	6	3.4	386	2	AD0098	probable sugar tra	373	6	3.4	464	2	C84686	probable AP2 domai
301	6	3.4	386	2	B97035	uncharacterized pr	374	6	3.4	467	2	T34874	hypothetical prote
302	6	3.4	387	2	A33624	spermidine/putresc	375	6	3.4	468	2	JC4285	acid phosphatase (
303	6	3.4	387	2	S41948	DNAJ protein - Str	376	6	3.4	469	2	JC5798	F0F1-ATPase (EC 3.
304	6	3.4	388	2	B75557	probable copropor	377	6	3.4	469	2	T48314	hypothetical prote
305	6	3.4	388	2	C83102	probable MFS trans	378	6	3.4	471	2	E83130	conserved hypotnet
306	6	3.4	390	2	E69272	conserved hypotnet	379	6	3.4	475	2	D90605	hexosephosphate tr
307	6	3.4	390	2	AC2761	penicillin binding	380	6	3.4	475	2	T18487	hypothetical prote
308	6	3.4	391	2	AG0533	probable drug effl	381	6	3.4	476	2	C72016	glycogen synthase
309	6	3.4	391	2	S21318	hypothetical prote	382	6	3.4	476	2	B86609	glycogen synthase
310	6	3.4	391	2	AH1762	hypothetical prote	383	6	3.4	478	2	D97682	hypothetical prote
311	6	3.4	391	2	G90253	formate hydrogenly	384	6	3.4	478	2	AD2907	cell division part
312	6	3.4	393	2	S74688	hypothetical prote	385	6	3.4	480	2	T18905	hypothetical prote
313	6	3.4	393	2	A64036	hypothetical prote	386	6	3.4	480	2	T18675	hypothetical prote
314	6	3.4	394	2	C85064	hypothetical prote	387	6	3.4	481	2	JC5378	protein disulfide-
315	6	3.4	394	2	S50696	hypothetical prote	388	6	3.4	483	2	T23312	hypothetical prote
316	6	3.4	396	1	C69988	hypothetical prote	389	6	3.4	484	2	F83005	thiazole biosynthe
317	6	3.4	397	2	B75592	chloramphenicol re	390	6	3.4	486	2	S43274	gag-like protein -
318	6	3.4	399	2	S70206	UDP-galactopyranos	391	6	3.4	487	2	S75338	NADH2 dehydrogenas
319	6	3.4	399	2	H87486	dnaJ protein - Str	392	6	3.4	488	2	I51379	synthrophin - Pacif
320	6	3.4	400	2	AH0666	1-deoxy-D-xylulose	393	6	3.4	490	2	T09084	phosphatidylinosit
321	6	3.4	400	2	H70658	probable aroF prot	394	6	3.4	491	2	E95082	lysyl-tRNA synthet

395	6	3.4	491	2	S73110	hypothetical prote	468	6	3.4	615	2	G90575	transketolase (tk)
396	6	3.4	493	2	S11148	amiA protein - Str	469	6	3.4	617	2	B46091	H-exporting ATPas
397	6	3.4	494	2	E86671	lysine-tRNA ligase	470	6	3.4	619	2	T44285	heat-shock-cognate
398	6	3.4	496	2	D86493	polymorphic outer	471	6	3.4	620	2	A70525	hypothetical prote
399	6	3.4	496	2	B97950	lysine-tRNA ligase	472	6	3.4	621	2	D71961	glucose inhibited
400	6	3.4	498	2	F98149	hypothetical prote	473	6	3.4	621	2	E64546	glucose inhibited
401	6	3.4	499	2	AH0485	probable insulin	474	6	3.4	621	2	A83421	peptidyl-prolyl ci
402	6	3.4	502	2	F90003	ATP synthase alpha	475	6	3.4	622	2	D44986	apical membrane an
403	6	3.4	503	2	AC3148	exopolysaccharide	476	6	3.4	622	2	B44986	apical membrane an
404	6	3.4	505	2	F90427	amino acid transpo	477	6	3.4	622	2	A32499	apical membrane an
405	6	3.4	506	2	C95217	choline transporte	478	6	3.4	622	2	A43866	apical membrane an
406	6	3.4	506	2	T44634	choline transporte	479	6	3.4	622	2	C44986	apical membrane an
407	6	3.4	506	2	C98081	hypothetical prote	480	6	3.4	627	2	G84558	probable SEC1 fami
408	6	3.4	509	2	T30861	traj protein - Sal	481	6	3.4	633	2	H84336	Hur3 transducer [1
409	6	3.4	510	2	S53958	hypothetical prote	482	6	3.4	633	2	G71026	hypothetical prote
410	6	3.4	512	2	T18803	hypothetical prote	483	6	3.4	633	2	F96652	protein P23N19.13
411	6	3.4	514	2	S18449	variant surface gl	484	6	3.4	634	2	T16897	hypothetical prote
412	6	3.4	514	2	S56384	hypothetical 50.3K	485	6	3.4	635	2	T46407	probable RNA helic
413	6	3.4	517	2	B71260	hypothetical prote	486	6	3.4	637	2	I49638	probable RNA helic
414	6	3.4	518	2	H86826	signal recognition	487	6	3.4	638	2	T28932	probable 4-coumara
415	6	3.4	520	2	AD2616	chromosomal replic	488	6	3.4	644	1	H69204	endopeptidase La h
416	6	3.4	520	2	S72324	orotate reductase	489	6	3.4	644	2	A86623	transketolase [imp
417	6	3.4	525	2	T00660	hypothetical prote	490	6	3.4	644	2	A72002	1-deoxyxylulose-5-
418	6	3.4	529	2	D97398	dnaA protein (L254	491	6	3.4	645	2	T11137	NADH2 dehydrogenas
419	6	3.4	529	2	A95423	probable ABC trans	492	6	3.4	645	2	G81315	membrane bound zin
420	6	3.4	532	2	G83424	hypothetical prote	493	6	3.4	653	2	F83405	probable TonB-depe
421	6	3.4	533	2	T04481	Mlo protein - barl	494	6	3.4	657	2	G86397	protein T7N9.18 [1
422	6	3.4	533	2	AE3138	hypothetical prote	495	6	3.4	658	1	A39018	carnitine O-palmit
423	6	3.4	533	2	T40153	hypothetical prote	496	6	3.4	658	2	A49362	carnitine O-palmit
424	6	3.4	535	2	A87570	ABC transporter, A	497	6	3.4	659	2	A95221	hypothetical prote
425	6	3.4	537	2	S39970	glucan 1,6-alpha-g	498	6	3.4	659	2	A98085	hypothetical prote
426	6	3.4	537	2	A38170	cytochrome d compl	499	6	3.4	662	2	A37226	glucose transport
427	6	3.4	540	2	H86777	hypothetical prote	500	6	3.4	663	2	C82415	ATP-dependent RNA
428	6	3.4	544	2	C86405	probable sphingos	501	6	3.4	668	2	G83349	conserved hypochet
429	6	3.4	547	2	S49815	transferrin-bindin	502	6	3.4	672	2	D81746	type III secretion
430	6	3.4	547	2	A44796	transferrin-bindin	503	6	3.4	676	2	A45900	complement C3b rec
431	6	3.4	548	2	A86385	hypothetical prote	504	6	3.4	676	2	AF1153	transcription anti
432	6	3.4	550	2	H98139	gumJ protein limpo	505	6	3.4	685	2	AE3408	exonuclease ABC c
433	6	3.4	551	2	JC7562	glioblastoma RING	506	6	3.4	687	1	B39045	protein-glutamine
434	6	3.4	551	2	T23945	hypothetical prote	507	6	3.4	690	2	A25534	hypothetical prote
435	6	3.4	556	1	A55483	transcription init	508	6	3.4	691	1	A29996	protein-glutamine
436	6	3.4	557	1	NUVKL	glucose-6-phosphat	509	6	3.4	691	2	G89935	penicillin-binding
437	6	3.4	557	2	D83478	gamma-glutamyltran	510	6	3.4	693	2	A45991	protein-glutamine
438	6	3.4	557	2	T45135	chaperone protein	511	6	3.4	693	2	B45991	protein-glutamine
439	6	3.4	558	2	C72391	conserved hypochet	512	6	3.4	698	1	S31714	prqN-glutamine
440	6	3.4	558	2	AE3580	iron(III)-transpor	513	6	3.4	698	1	S31630	NAD+-protein APP-r
441	6	3.4	561	2	T01950	hypothetical prote	514	6	3.4	698	2	F83096	probable two-compo
442	6	3.4	562	2	AH2084	hypothetical prote	515	6	3.4	698	2	E82443	heme transport pro
443	6	3.4	563	2	S32156	mandelonitrile lya	516	6	3.4	702	2	H84189	lipotease protein li
444	6	3.4	568	1	D64059	2-succinyl-6-hydro	517	6	3.4	708	2	I83196	NEDD-4 ORF - mouse
445	6	3.4	570	2	H90370	thermopsin precurs	518	6	3.4	708	2	AH2626	ferrichrome iron r
446	6	3.4	571	2	T10232	hypothetical prote	519	6	3.4	713	2	JC2534	RVIG protein - rat
447	6	3.4	572	2	C75515	conserved hypochet	520	6	3.4	721	2	C84073	DNA topoisomerase
448	6	3.4	572	2	T08509	trbL protein - Ent	521	6	3.4	722	2	F71270	probable flagellar
449	6	3.4	574	2	G69769	pyruvate oxidase h	522	6	3.4	724	2	T27858	probable peroxidase
450	6	3.4	575	1	S35138	acetolactate synth	523	6	3.4	724	2	S57604	probable membrane
451	6	3.4	577	2	D83473	isocitrate dehydro	524	6	3.4	728	2	F97262	topoisomerase B [1
452	6	3.4	582	2	H81671	single-stranded-DN	525	6	3.4	731	1	A55800	cucumislin (EC 3.4.
453	6	3.4	582	2	C59099	hypothetical prote	526	6	3.4	738	2	T00343	hypothetical prote
454	6	3.4	582	2	A55174	kinase-associated	527	6	3.4	739	2	T10648	hypothetical prote
455	6	3.4	583	2	H84810	probable calcium-d	528	6	3.4	744	2	C75276	hypothetical prote
456	6	3.4	584	2	H71513	probable ssdna exo	529	6	3.4	747	2	G97408	sensory box sensor
457	6	3.4	585	2	S48929	hypothetical prote	530	6	3.4	751	2	T01449	regA protein U6140
458	6	3.4	587	2	A31776	lactose permease -	531	6	3.4	759	2	T24130	cytoskeletal prote
459	6	3.4	595	2	T06111	hypothetical prote	532	6	3.4	759	2	B70585	hypothetical prote
460	6	3.4	596	1	A44072	transcription fact	533	6	3.4	763	2	A43747	X-Pro dipeptidyl-p
461	6	3.4	596	2	T40609	DNA primase (EC 2.	534	6	3.4	763	2	A43748	X-Pro dipeptidyl-p
462	6	3.4	596	2	C97060	DNA primase, DNAG	535	6	3.4	763	2	A86881	hypothetical prote
463	6	3.4	598	2	I39596	chitinase - Alter	536	6	3.4	763	2	T22843	hypothetical prote
464	6	3.4	607	2	S48326	hypothetical prote	537	6	3.4	767	2	S63220	probable membrane
465	6	3.4	610	2	C96732	hypothetical prote	538	6	3.4	780	2	T30184	hypothetical prote
466	6	3.4	611	2	A85000	heat shock protein	539	6	3.4	787	2	T16901	hypothetical prote
467	6	3.4	614	1	S61532	RET oncogene fusio	540	6	3.4	787	2	S56268	hypothetical prote

541	6	3.4	791	2	F81056	hemoglobin recepto	614	3.4	1251	2	JH0256	botulinum neurotox	
542	6	3.4	792	2	S61335	hemoglobin recepto	615	3.4	1252	2	S21178	botulinum neurotox	
543	6	3.4	793	1	KXRTF	furin (EC 3.4.21.7	616	3.4	1286	1	RJBOP	interphotoreceptor	
544	6	3.4	796	1	A32434	abaA protein - Eme	617	3.4	1290	2	S73653	DNA-directed RNA p	
545	6	3.4	803	2	AB0530	outer membrane pro	618	3.4	1297	2	S37991	neurotoxin - Clost	
546	6	3.4	805	2	T48840	transducer protein	619	3.4	1301	2	D85188	disease resistance	
547	6	3.4	806	2	E64221	phenylalanine-tRNA	620	3.4	1305	2	T12897	DNA polymerase III	
548	6	3.4	806	2	T08932	hypothetical prote	621	3.4	1338	2	A49634	aldehyde oxidase (	
549	6	3.4	809	2	T41645	probable spindle p	622	3.4	1341	2	H98323	hypothetical prote	
550	6	3.4	811	2	B82365	glycerol-3-phospha	623	3.4	1351	2	E97273	superfamily I DNA	
551	6	3.4	812	1	MMECOF	outer membrane u	624	3.4	1355	2	T00961	hypothetical prote	
552	6	3.4	815	2	AB2444	hypothetical prote	625	3.4	1375	2	JT0345	dextranucrase (EC	
553	6	3.4	819	2	A13197	TonB-dependent rec	626	3.4	1400	2	T31555	hypothetical prote	
554	6	3.4	822	2	T33163	hypothetical prote	627	3.4	1430	2	AF0351	probable autotrans	
555	6	3.4	826	2	D86458	hypothetical prote	628	3.4	1446	2	T13018	hypothetical prote	
556	6	3.4	828	2	T03544	hypothetical prote	629	3.4	1449	2	B84426	hypothetical prote	
557	6	3.4	833	1	S20387	outer membrane pro	630	3.4	1452	2	A97323	DNA polymerase III	
558	6	3.4	833	2	T14703	F1 capsule anchori	631	3.4	1470	2	S45323	genome polyprotein	
559	6	3.4	839	2	B96576	hypothetical prote	632	3.4	1473	2	T31422	C-terminal domain-	
560	6	3.4	843	2	S78372	DNA-directed RNA p	633	3.4	1475	2	B33135	gtfB protein precu	
561	6	3.4	854	2	T23155	hypothetical prote	634	3.4	1476	1	A39901	cystic fibrosis tr	
562	6	3.4	858	2	A44919	GCR3 protein - yea	635	3.4	1476	1	A40303	cystic fibrosis tr	
563	6	3.4	859	2	G86242	hypothetical prote	636	3.4	1476	2	JC5143	alpha-macroglobuli	
564	6	3.4	860	2	T28227	ORF MSV067 probabl	637	3.4	1514	2	T52080	multi resistance p	
565	6	3.4	870	2	AD0625	aminopeptidase N I	638	3.4	1532	2	T18438	hypothetical prote	
566	6	3.4	873	2	S75028	hypothetical prote	639	3.4	1597	1	BVFESL	sol protein, large	
567	6	3.4	875	2	T17382	vric protein - Dic	640	3.4	1597	2	T08428	gene small optic l	
568	6	3.4	875	2	T27495	hypothetical prote	641	3.4	1638	2	T25352	hypothetical prote	
569	6	3.4	887	2	S70642	ubiquitin ligase N	642	3.4	1652	2	T16799	hypothetical prote	
570	6	3.4	903	2	D87250	DNA mismatch repai	643	3.4	1679	2	S49802	probable membrane	
571	6	3.4	905	2	T23229	hypothetical prote	644	3.4	1755	2	T51532	hypothetical prote	
572	6	3.4	906	1	RNBPK1	DNA-directed RNA p	645	3.4	1762	2	T03222	probable polyketid	
573	6	3.4	908	2	JN0819	transferrin-bindin	646	3.4	1817	2	T10689	hypothetical prote	
574	6	3.4	916	2	H69161	DNA helicase II -	647	3.4	1937	2	T03224	probable polyketid	
575	6	3.4	919	1	QVEFG	phosphoenolpyruvat	648	3.4	2051	2	S34688	enoyl-(acyl)-carrie	
576	6	3.4	936	2	C86546	polymorphic outer	649	3.4	2073	2	T43311	fatty acyl-CoA syn	
577	6	3.4	936	2	B81591	polymorphic membra	650	3.4	2073	2	T3207	fatty acid synthas	
578	6	3.4	936	2	C72078	polymorphic outer	651	3.4	2098	2	T13166	rough deal protein	
579	6	3.4	940	2	H71409	hypothetical prote	652	3.4	2100	2	T03223	probable polyketid	
580	6	3.4	946	2	F84280	ATP-dependent heli	653	3.4	2133	2	T42763	coagulation factor	
581	6	3.4	953	1	B30169	Leukotoxin A - Pas	654	3.4	2149	2	C96695	ribulose bisphosph	
582	6	3.4	954	2	A10438	probable exported	655	3.4	2214	2	T16305	hypothetical prote	
583	6	3.4	960	2	JE0356	gamma-aminobutylic	656	3.4	2319	2	A47004	coagulation factor	
584	6	3.4	982	2	T34830	polyketide synthas	657	3.4	2351	1	EZHU	coagulation factor	
585	6	3.4	988	2	T51054	related to alpha-a	658	3.4	3014	1	JC5620	genome polyprotein	
586	6	3.4	992	1	GNWVR3	structural polypro	659	3.4	3078	2	T28432	variant-specific s	
587	6	3.4	992	2	JQ1165	Env protein - Maed	660	3.4	3268	2	S69625	hypothetical prote	
588	6	3.4	996	2	E98092	cylM protein, cyto	661	3.4	3433	1	S28381	utrophin - human	
589	6	3.4	996	2	A95228	bacteriocin format	662	3.4	3512	2	T17121	CPY protein - midg	
590	6	3.4	1004	2	JC2221	major surface glyco	663	3.4	4092	1	S38128	dynein heavy chain	
591	6	3.4	1014	1	NMCLSS	exo-alpha-sialidas	664	3.4	4351	2	T00252	MEGF1 protein - ra	
592	6	3.4	1018	2	A54744	contactin 1 precu	665	3.4	4436	2	E71086	hypothetical prote	
593	6	3.4	1032	2	T40658	serine-threonine p	666	3.4	4589	2	T14914	dynein beta heavy	
594	6	3.4	1037	2	T27345	hypothetical prote	667	3.4	4613	2	T17409	polyketide synthas	
595	6	3.4	1038	2	S38801	pullulanase (EC 3.	668	3.4	4639	1	A54794	dynein heavy chain	
596	6	3.4	1049	2	C95883	probable efflux pr	669	3.4	4735	2	T17463	rifamycin polyketi	
597	6	3.4	1052	2	AF2959	conserved hypothet	670	3.4	4981	2	T18489	hypothetical prote	
598	6	3.4	1054	2	T43226	translation initia	671	3.4	6831	2	A88852	protein unc-22 (lm	
599	6	3.4	1058	2	AG2541	cation efflux syst	672	3.4	6839	2	S57242	twichin [similari	
600	6	3.4	1063	1	GNWV77	structural polypro	673	3.4	7160	2	T27935	hypothetical prote	
601	6	3.4	1078	2	T19745	hypothetical prote	674	3.4	8563	2	T30226	polyketide synthas	
602	6	3.4	1090	2	S11823	pullulanase (EC 3.	675	3.4	10223	2	T30225	polyketide synthas	
603	6	3.4	1096	2	A26879	pullulanase (EC 3.	676	5	2.8	13	2	S36874	cytochrome P450 CM
604	6	3.4	1118	2	S48383	probable membrane	677	5	2.8	14	2	A47421	leukotriene B-4 12
605	6	3.4	1154	2	T18525	diacylglycerol kin	678	5	2.8	18	2	A43826	dnak-type molecula
606	6	3.4	1173	2	T31421	C-terminal domain-	679	5	2.8	20	2	S21737	comarin 7-monooxy
607	6	3.4	1184	2	H71436	hypothetical prote	680	5	2.8	20	2	S02002	photosystem II pro
608	6	3.4	1195	2	S61886	chitin synthase (E	681	5	2.8	20	2	PN0459	basic eosinophil p
609	6	3.4	1217	2	H89863	hypothetical prote	682	5	2.8	21	2	S03979	testosterone 15alp
610	6	3.4	1228	2	C98219	proline dehydrogen	683	5	2.8	25	2	JA0171	dnak-type molecula
611	6	3.4	1234	2	AG3067	proline dehydrogen	684	5	2.8	28	2	I45911	dnak-type molecula
612	6	3.4	1234	2	C97606	hypothetical prote	685	5	2.8	28	2	A41763	extracellular prot
613	6	3.4	1234	2	AE2828	conserved hypothet	686	5	2.8	30	2	S12172	hypothetical prote

687	5	2.8	30	2	B84053	hypothetical prote	760	5	2.8	69	2	S45610	guanylate cyclase
688	5	2.8	33	2	P00150	dnak-type molecula	761	5	2.8	70	2	F84752	hypothetical prote
689	5	2.8	33	2	S34505	hypothetical prote	762	5	2.8	70	2	S43419	conserved hypotet
690	5	2.8	34	2	I50713	homeobox protein -	763	5	2.8	71	2	T43815	conserved hypotet
691	5	2.8	34	2	I51364	homeobox protein -	764	5	2.8	71	2	JH0349	T-cell receptor be
692	5	2.8	34	2	I38900	homeobox protein -	765	5	2.8	71	2	S21364	dnak-type molecula
693	5	2.8	34	2	I49145	homeobox protein -	766	5	2.8	71	2	T39627	hypothetical prote
694	5	2.8	35	4	S41911	hypothetical prote	767	5	2.8	71	2	I70114	arrestin-D - rat (
695	5	2.8	36	1	F2RZ1	photosystem II pro	768	5	2.8	72	2	C30517	Ig heavy chain pre
696	5	2.8	36	1	F2NTI	photosystem II pro	769	5	2.8	72	2	A71970	cbh3-type cytochro
697	5	2.8	36	1	F2KHD	photosystem II pro	770	5	2.8	72	2	C83219	hypothetical prote
698	5	2.8	36	1	I59139	dnak-type molecula	771	5	2.8	72	2	E81126	conserved hypotet
699	5	2.8	36	2	S60133	photosystem II pro	772	5	2.8	72	2	G85909	partial probable t
700	5	2.8	36	2	S01044	photosystem II pro	773	5	2.8	72	2	AE3444	hypothetical prote
701	5	2.8	36	2	S28767	photosystem II pro	774	5	2.8	73	2	F90221	lsu ribosomal prot
702	5	2.8	36	2	S07877	photosystem II pro	775	5	2.8	73	2	B64538	cbh3-type cytochro
703	5	2.8	36	2	A05019	photosystem II pro	776	5	2.8	74	1	CBK75E	cytochrome b559 co
704	5	2.8	36	2	JN0315	photosystem II pro	777	5	2.8	74	2	C96547	hypothetical prote
705	5	2.8	36	2	S58535	photosystem II pro	778	5	2.8	74	2	C81197	hypothetical prote
706	5	2.8	39	2	S75180	photosystem II pro	779	5	2.8	74	2	AH3104	conserved hypotet
707	5	2.8	39	2	A42286	photosystem II pro	780	5	2.8	75	2	A81297	hypothetical prote
708	5	2.8	40	2	A0128	probable antigen 1	781	5	2.8	75	2	F70761	hypothetical prote
709	5	2.8	40	2	B97413	hypothetical prote	782	5	2.8	75	2	AF2565	hypothetical prote
710	5	2.8	41	2	E81565	hypothetical prote	783	5	2.8	76	2	D81094	hypothetical prote
711	5	2.8	43	2	S41388	protein 3a - human	784	5	2.8	76	2	C82511	hypothetical prote
712	5	2.8	44	2	P00181	naringenin-chalcon	785	5	2.8	76	2	G82739	hypothetical prote
713	5	2.8	44	2	H69676	phosphatase (Rapa)	786	5	2.8	77	2	G83542	hypothetical prote
714	5	2.8	44	2	S21028	m-factor precursor	787	5	2.8	77	2	T19575	epoxide hydrolase
715	5	2.8	46	2	A10624	probable DNA inver	788	5	2.8	78	2	S17801	hypothetical prote
716	5	2.8	47	2	F81118	hypothetical prote	789	5	2.8	78	2	T08483	kleA protein - Ent
717	5	2.8	48	2	D82534	hypothetical prote	790	5	2.8	78	2	S72611	hypothetical prote
718	5	2.8	49	2	JH0843	GTP-binding protel	791	5	2.8	78	2	S72971	hypothetical prote
719	5	2.8	49	2	G08452	hypothetical prote	792	5	2.8	78	2	C98182	hypothetical prote
720	5	2.8	50	2	S08756	hypothetical prote	793	5	2.8	79	2	F95210	conserved domain p
721	5	2.8	50	2	AG2783	hypothetical prote	794	5	2.8	80	2	C91287	hypothetical prote
722	5	2.8	51	2	F82759	hypothetical prote	795	5	2.8	80	2	B83097	hypothetical prote
723	5	2.8	51	2	E95849	hypothetical prote	796	5	2.8	80	2	A54059	protein kinase C (
724	5	2.8	52	2	T07434	photosystem II pro	797	5	2.8	80	2	A10707	cell division acti
725	5	2.8	52	2	B90462	hypothetical prote	798	5	2.8	81	1	CBY855	cytochrome b559 co
726	5	2.8	52	2	T07269	hypothetical prote	799	5	2.8	81	2	F85101	hypothetical prote
727	5	2.8	54	2	S35697	leukocidin chain F	800	5	2.8	81	2	H70237	hypothetical prote
728	5	2.8	55	2	T43598	hypothetical prote	801	5	2.8	81	2	C83792	hypothetical prote
729	5	2.8	55	2	F56976	transfer complex p	802	5	2.8	82	1	S53882	cytochrome b559 co
730	5	2.8	55	2	A70070	hypothetical prote	803	5	2.8	82	2	E84301	hypothetical prote
731	5	2.8	55	2	AC2366	hypothetical prote	804	5	2.8	82	2	S39706	hypothetical prote
732	5	2.8	56	2	S65513	ribulose-bisphosph	805	5	2.8	82	2	B97702	hypothetical prote
733	5	2.8	56	2	S74630	hypothetical prote	806	5	2.8	82	2	AF2286	cytochrome b559 al
734	5	2.8	56	2	C69044	hypothetical prote	807	5	2.8	83	1	CBMT55	cytochrome b559 co
735	5	2.8	56	2	H98056	hypothetical prote	808	5	2.8	83	1	S00418	cytochrome b559 co
736	5	2.8	57	1	B64240	ribosomal protein	809	5	2.8	83	1	CBMT55	cytochrome b559 co
737	5	2.8	57	2	S26205	metallothionein -	810	5	2.8	83	1	CBMT5E	cytochrome b559 co
738	5	2.8	57	2	AC1756	hypothetical prote	811	5	2.8	83	1	A29956	cytochrome b559 co
739	5	2.8	58	2	I77464	luteinizing hormon	812	5	2.8	83	1	S03191	cytochrome b559 co
740	5	2.8	58	2	AD2652	hypothetical prote	813	5	2.8	83	1	S58568	cytochrome b559 co
741	5	2.8	59	2	B27606	Ig heavy chain V-a	814	5	2.8	83	1	A48310	cytochrome b559 co
742	5	2.8	60	2	B90490	hypothetical prote	815	5	2.8	83	1	S55789	cytochrome b559 co
743	5	2.8	60	2	T28340	ORF MSV178 hypothe	816	5	2.8	83	1	S01243	cytochrome b559 co
744	5	2.8	60	2	H95354	protein [imported	817	5	2.8	83	1	CBLV55	cytochrome b559 co
745	5	2.8	61	2	PN0561	rod outer segment	818	5	2.8	83	1	IKKREG	high potential iro
746	5	2.8	61	2	D83274	conserved hypotet	819	5	2.8	83	2	T07257	cytochrome b559 co
747	5	2.8	64	2	S78309	hypothetical prote	820	5	2.8	83	2	T07475	cytochrome b559 co
748	5	2.8	64	2	S51273	hypothetical prote	821	5	2.8	83	2	B84174	hypothetical prote
749	5	2.8	64	2	T34792	probable transcrip	822	5	2.8	84	2	S78337	cytochrome b559 co
750	5	2.8	64	2	S23540	hypothetical prote	823	5	2.8	84	2	S73312	cytochrome b559 al
751	5	2.8	65	2	C46448	hypothetical prote	824	5	2.8	84	2	S21365	dnak-type molecula
752	5	2.8	65	2	G81292	hypothetical prote	825	5	2.8	84	2	S33999	EP84R protein - Af
753	5	2.8	66	2	H81804	hypothetical prote	826	5	2.8	84	2	T18097	hypothetical prote
754	5	2.8	66	2	B81061	hypothetical prote	827	5	2.8	85	1	R3NT16	ribosomal protein
755	5	2.8	67	2	H90319	hypothetical prote	828	5	2.8	85	2	F75291	hypothetical prote
756	5	2.8	67	2	I54475	HLA-DNA-related sm	829	5	2.8	86	2	H90682	hypothetical prote
757	5	2.8	68	2	D95025	ribosomal protein	830	5	2.8	86	2	F64527	hypothetical prote
758	5	2.8	68	2	E97896	50S ribosomal prot	831	5	2.8	86	2	E71978	hypothetical prote
759	5	2.8	68	2	B03309	dnak-type molecula	832	5	2.8	86	2	C81000	hypothetical prote



833	5	2.8	86	2	S66089	conserved hypotet	906	5	2.8	99	2	AI2403	hypothetical prote
834	5	2.8	86	2	A35241	IgE Fc receptor ga	907	5	2.8	100	2	T01656	leucyl-cRNA synthe
835	5	2.8	86	2	A46461	T-cell receptor et	908	5	2.8	100	2	JX0302	ubiquinol-cytochro
836	5	2.8	86	2	S02118	IgE Fc receptor I.	909	5	2.8	100	2	AB2710	conserved hypotet
837	5	2.8	86	2	D85533	hypothetical prote	910	5	2.8	100	2	C97492	hypothetical prote
838	5	2.8	86	2	AG0130	conserved hypotet	911	5	2.8	100	2	B43998	hypothetical prote
839	5	2.8	86	2	F64766	yaib protein - Esc	912	5	2.8	100	2	D84081	hypothetical prote
840	5	2.8	87	2	S77861	ribosomal protein	913	5	2.8	100	2	AE2669	conserved hypotet
841	5	2.8	87	2	B86743	conserved hypotet	914	5	2.8	101	2	C31982	Ca2+-transporting
842	5	2.8	87	2	E90933	Cell division acti	915	5	2.8	101	2	JH0354	T-cell receptor be
843	5	2.8	87	2	A85782	Cell division acti	916	5	2.8	101	2	JC7579	ribosomal protein
844	5	2.8	87	2	C64932	Cell division acti	917	5	2.8	101	2	A29165	proteoglycan link
845	5	2.8	88	1	QQ8CC8	hypothetical prote	918	5	2.8	101	2	A05124	hypothetical prote
846	5	2.8	88	2	T07057	ribosomal protein	919	5	2.8	101	2	D75013	hypothetical prote
847	5	2.8	88	2	G90517	hypothetical prote	920	5	2.8	101	2	H97263	PTS system IIB com
848	5	2.8	88	2	A84166	hypothetical prote	921	5	2.8	101	2	AD0405	conserved hypotet
849	5	2.8	88	2	C82468	hypothetical prote	922	5	2.8	102	2	JA0170	dnak-type molecula
850	5	2.8	88	2	C83823	hypothetical prote	923	5	2.8	102	2	AF1661	phosphotransferase
851	5	2.8	89	2	A90339	hypothetical prote	924	5	2.8	102	2	AH1289	phosphotransferase
852	5	2.8	89	2	AC1576	carbon dioxide con	925	5	2.8	102	2	B24444	hypothetical rRNA
853	5	2.8	89	2	AH1222	carbon dioxide con	926	5	2.8	102	2	S28361	hypothetical prote
854	5	2.8	89	4	T46004	probable pseudogen	927	5	2.8	102	2	S69463	hypothetical prote
855	5	2.8	90	2	S05375	dnak-type molecula	928	5	2.8	102	2	D88428	protein R07E5.12 l
856	5	2.8	90	2	B91003	probable regulator	929	5	2.8	102	2	S44781	C30A5.4 protein -
857	5	2.8	90	2	G84991	hypothetical prote	930	5	2.8	102	2	A72739	probable DNA-direc
858	5	2.8	90	2	C83643	hypothetical prote	931	5	2.8	103	2	AD1003	conserved hypotet
859	5	2.8	90	2	A81245	hypothetical prote	932	5	2.8	103	2	G83241	conserved hypotet
860	5	2.8	90	2	H82023	probable lipoprote	933	5	2.8	103	2	I37384	FAS soluble protei
861	5	2.8	90	2	C69480	hypothetical prote	934	5	2.8	103	2	AG1563	hypothetical prote
862	5	2.8	90	2	H85847	hypothetical prote	935	5	2.8	103	2	AH1206	hypothetical prote
863	5	2.8	91	1	RGBBHD	mercuric resistanc	936	5	2.8	103	2	E75071	hypothetical prote
864	5	2.8	91	1	S09524	mercuric resistanc	937	5	2.8	104	1	CCFG	cytochrome c - bul
865	5	2.8	91	1	RGPSHA	mercuric resistanc	938	5	2.8	104	2	JN0643	naphthalene 1,2-di
866	5	2.8	91	2	I54404	T cell antigen rec	939	5	2.8	104	2	S04719	ribosomal protein
867	5	2.8	91	2	S07895	hypothetical prote	940	5	2.8	104	2	G42721	recombination prot
868	5	2.8	91	2	JC7393	medaka-type gonado	941	5	2.8	104	2	S45400	probable membrane
869	5	2.8	91	2	S72298	hypothetical prote	942	5	2.8	104	2	A72649	hypothetical prote
870	5	2.8	92	2	A82141	integration host f	943	5	2.8	104	2	AE1430	B. subtilis yuid p
871	5	2.8	92	2	A84397	50S ribosomal prot	944	5	2.8	104	2	S00562	homeotic protein E
872	5	2.8	92	2	F69390	LSU ribosomal prot	945	5	2.8	104	2	AE2131	hypothetical prote
873	5	2.8	92	2	PL0223	T-cell receptor be	946	5	2.8	105	2	I51851	parathyroid hormon
874	5	2.8	93	2	PL0224	T-cell receptor be	947	5	2.8	105	2	T10227	hypothetical prote
875	5	2.8	94	1	BCBU1A	S-100 protein alph	948	5	2.8	105	2	S70042	hypothetical prote
876	5	2.8	94	1	BCBO1A	S-100 protein alph	949	5	2.8	105	2	AB3256	hypothetical cytos
877	5	2.8	94	2	S03380	major fecal allergy	950	5	2.8	106	1	R8BY2B	acidic ribosomal p
878	5	2.8	94	2	A49832	DNA gyrase A - Sta	951	5	2.8	106	2	T16378	hypothetical prote
879	5	2.8	94	2	B49828	T-cell receptor be	952	5	2.8	106	2	T36973	hypothetical prote
880	5	2.8	94	2	C49828	T-cell receptor be	953	5	2.8	107	2	JH0352	T-cell receptor be
881	5	2.8	94	2	F86566	CT490 hypothetical	954	5	2.8	107	2	S36262	Ig lambda chain v
882	5	2.8	94	2	T01217	hypothetical prote	955	5	2.8	107	2	A69547	conserved hypotet
883	5	2.8	94	2	C72058	conserved hypotet	956	5	2.8	107	2	H82729	hypothetical prote
884	5	2.8	94	2	D71261	hypothetical prote	957	5	2.8	107	2	AI0282	probable DNA-bindi
885	5	2.8	95	2	A49027	T-cell receptor va	958	5	2.8	107	4	A60500	hypothetical mutan
886	5	2.8	95	2	H69767	hypothetical prote	959	5	2.8	108	2	S36277	Ig lambda chain v
887	5	2.8	95	2	AC3625	hypothetical prote	960	5	2.8	108	2	S32883	Ig lambda chain v
888	5	2.8	96	2	D46289	keratinocyte growt	961	5	2.8	108	2	S22892	T-cell receptor al
889	5	2.8	96	2	AF0741	conserved hypotet	962	5	2.8	108	2	S23204	retinol-binding pr
890	5	2.8	96	2	B90793	hypothetical prote	963	5	2.8	108	2	S83765	hypothetical prote
891	5	2.8	96	2	T31209	hypothetical prote	964	5	2.8	108	2	AF0116	conserved hypotet
892	5	2.8	96	2	G85653	hypothetical prote	965	5	2.8	108	2	AE0879	conserved hypotet
893	5	2.8	96	2	D95949	hypothetical prote	966	5	2.8	108	2	T33603	hypothetical prote
894	5	2.8	96	2	AF3175	attachment protein	967	5	2.8	109	2	H84139	PTS system, cellob
895	5	2.8	97	2	JT0703	ferredoxin-thiorel	968	5	2.8	109	2	S58825	hypothetical prote
896	5	2.8	97	2	B86345	F16P4.12 protein -	969	5	2.8	109	2	D70077	hypothetical prote
897	5	2.8	97	2	S03497	T-cell receptor be	970	5	2.8	109	2	AE0549	hypothetical prote
898	5	2.8	98	2	S17390	T-cell receptor be	971	5	2.8	109	2	S31950	penicillin-binding
899	5	2.8	98	2	PH0874	Ig heavy chain v r	972	5	2.8	110	1	R5BYA1	acidic ribosomal p
900	5	2.8	98	2	D96797	Sm-like protein [i	973	5	2.8	110	2	AG3382	quaternary ammoni
901	5	2.8	99	2	C60339	spap surface antig	974	5	2.8	110	2	A69902	hypothetical prote
902	5	2.8	99	2	S07268	regulatory protein	975	5	2.8	110	2	E71004	hypothetical prote
903	5	2.8	99	2	A70608	probable PE protei	976	5	2.8	110	2	AI2898	conserved hypotet
904	5	2.8	99	2	F86735	hypothetical prote	977	5	2.8	111	2	G70521	hypothetical prote
905	5	2.8	99	2	F82709	hypothetical prote	978	5	2.8	111	2	T31580	hypothetical prote

979 5 2.8 111 2 AD1313 hypothetical prote  
 980 5 2.8 111 2 A75322 hypothetical prote  
 981 5 2.8 112 2 JU0151 biphenyl-2,3-diol  
 982 5 2.8 112 2 S25572 Ig heavy chain V r  
 983 5 2.8 112 2 S69318 hypothetical prote  
 984 5 2.8 112 2 H71118 hypothetical prote  
 985 5 2.8 112 2 T47738 hypothetical prote  
 986 5 2.8 112 2 G70855 hypothetical prote  
 987 5 2.8 112 2 A87239 pyridine transhydr  
 988 5 2.8 112 2 AE0496 probable lipoprote  
 989 5 2.8 113 1 CCCRCE cytochrome c - Cri  
 990 5 2.8 113 1 R6DOP1 acidic ribosomal p  
 991 5 2.8 113 2 JH0351 T-cell receptor be  
 992 5 2.8 113 2 I38312 T-cell receptor be  
 993 5 2.8 113 2 E27664 T-cell receptor be  
 994 5 2.8 113 2 D27664 T-cell receptor be  
 995 5 2.8 113 2 D27664 T-cell receptor be  
 996 5 2.8 113 2 C70927 probable ribosomal  
 997 5 2.8 113 2 G87110 50S ribosomal prot  
 998 5 2.8 113 2 B40899 hypothetical prote  
 999 5 2.8 113 2 T44302 hypothetical prote  
 1000 5 2.8 113 2 C71074 hypothetical prote

## ALIGNMENTS

## RESULT 1

outer membrane protein P1 precursor, subtype 6U - Haemophilus influenzae (strain 8358, t  
 C:Species: Haemophilus influenzae  
 C:Date: 28-Oct-1992 #sequence\_revision 30-Jan-1993 #text\_change 12-Dec-1997  
 C:Accession: B40183  
 R:Munson Jr., R.; Grass, S.; Einhorn, M.; Bailey, C.; Newell, C.  
 Infect. Immun. 57, 3300-3305, 1989  
 A:Title: Comparative analysis of the structures of the outer membrane protein P1 genes f  
 A:Reference number: A40183; MUID:90035394; PMID:2572549  
 A:Accession: B40183  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-450 <MUN>  
 A:Cross-references: GB:M27683  
 C:Superfamily: long-chain fatty acid transport protein fadL  
 C:Keywords: membrane protein  
 F:1-22/Domain: signal sequence #status predicted <SIG>  
 F:23-450/Product: outer membrane protein P1 #status predicted <MAT>

Query Match 7.3%; Score 13; DB 2; Length 450;  
 Best Local Similarity 100.0%; Pred. No. 3.7e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 GATYKFTPNLSVD 133  
 |||  
 Db 391 GATYKFTPNLSVD 403

## RESULT 2

outer membrane protein P1 precursor, subtype 3L - Haemophilus influenzae (strain 1613, t  
 C:Species: Haemophilus influenzae  
 C:Date: 28-Oct-1992 #sequence\_revision 30-Jan-1993 #text\_change 20-Aug-1999  
 C:Accession: A40183  
 R:Munson Jr., R.; Grass, S.; Einhorn, M.; Bailey, C.; Newell, C.  
 Infect. Immun. 57, 3300-3305, 1989  
 A:Title: Comparative analysis of the structures of the outer membrane protein P1 genes f  
 A:Reference number: A40183; MUID:90035394; PMID:2572549  
 A:Accession: A40183  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-455 <MUN>  
 A:Cross-references: GB:M63151; GB:M27682; NID:g148954; PIDN:AAA24991.1; PID:g148955  
 C:Superfamily: long-chain fatty acid transport protein fadL  
 C:Keywords: membrane protein

Query Match 7.3%; Score 13; DB 2; Length 459;  
 Best Local Similarity 100.0%; Pred. No. 3.7e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 GATYKFTPNLSVD 133  
 |||  
 Db 397 GATYKFTPNLSVD 409

## RESULT 5

F:1-22/Domain: signal sequence #status predicted <SIG>  
 F:23-455/Product: outer membrane protein P1 #status predicted <MAT>

Query Match 7.3%; Score 13; DB 2; Length 455;  
 Best Local Similarity 100.0%; Pred. No. 3.7e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 GATYKFTPNLSVD 133  
 |||  
 Db 393 GATYKFTPNLSVD 405

## RESULT 3

outer membrane protein P1 precursor - Haemophilus influenzae (type b)  
 C:Species: Haemophilus influenzae  
 C:Date: 19-Nov-1988 #sequence\_revision 19-Nov-1988 #text\_change 20-Aug-1999  
 C:Accession: A30510; A28787  
 R:Munson Jr., R.; Grass, S.  
 Infect. Immun. 56, 2235-2242, 1988  
 A:Title: Purification, cloning, and sequence of outer membrane protein P1 of Haemophi  
 A:Reference number: A30510; MUID:88314258; PMID:2842261  
 A:Accession: A30510  
 A:Molecule type: DNA  
 A:Residues: 1-459 <MUN>  
 A:Cross-references: GB:J03381; NID:g148952; PIDN:AAA24990.1; PID:g148953  
 C:Superfamily: long-chain fatty acid transport protein fadL  
 C:Keywords: membrane protein  
 F:1-22/Domain: signal sequence #status predicted <SIG>  
 F:23-459/Product: outer membrane protein P1 #status predicted <MAT>

Query Match 7.3%; Score 13; DB 2; Length 459;  
 Best Local Similarity 100.0%; Pred. No. 3.7e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 GATYKFTPNLSVD 133  
 |||  
 Db 397 GATYKFTPNLSVD 409

## RESULT 4

outer membrane protein P1 precursor - Haemophilus influenzae (strain Rd KW20)  
 C:Species: Haemophilus influenzae  
 C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 20-Aug-1999  
 C:Accession: F64065  
 R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage  
 ; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman  
 ; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.  
 Science 269, 496-512, 1995  
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter  
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
 A:Reference number: A64000; MUID:95350630; PMID:7542800  
 A:Accession: F64065  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-459 <TIGR>

A:Cross-references: GB:U32723; GB:I42023; NID:g1573363; PIDN:AAC22060.1; PID:g1573372  
 C:Superfamily: long-chain fatty acid transport protein fadL  
 C:Keywords: membrane protein  
 F:1-22/Domain: signal sequence #status predicted <SIG>  
 F:23-459/Product: outer membrane protein P1 #status predicted <MAT>

Query Match 7.3%; Score 13; DB 2; Length 459;  
 Best Local Similarity 100.0%; Pred. No. 3.7e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 GATYKFTPNLSVD 133  
 |||  
 Db 397 GATYKFTPNLSVD 409

G82406  
 long-chain fatty acid transport protein VCA0862 [imported] - Vibrio cholerae (strain N16)  
 C:Species: Vibrio cholerae  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 C:Accession: G82406  
 R:Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
 Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, H.  
 I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
 A:Reference number: A82035; MUID:20406833; PMID:10952301  
 A:Accession: G82406  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-412 <HEI>  
 A:Cross-references: GB:AE004414; GB:AE003853; NID:g9658293; PIDN:AAF96760.1; GSPDB:GN001  
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C:Genetics:  
 A:Gene: VCA0862  
 A:Map position: 2  
 C:Superfamily: long-chain fatty acid transport protein fadL

Query Match 5.0%; Score 9; DB 2; Length 412;  
 Best Local Similarity 100.0%; Pred. No. 0.45;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 WELSGFHQL 31  
 |||||  
 DB 261 WELSGFHQL 269

RESULT 6  
 H69448  
 heme d1 biosynthesis protein (nirD) homolog - Archaeoglobus fulgidus  
 C:Species: Archaeoglobus fulgidus  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
 C:Accession: H69448  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
 ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Uitterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.  
 Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
 A:Reference number: A69250; MUID:98049343; PMID:9389475  
 A:Accession: H69448  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-72 <KLE>  
 A:Cross-references: GB:AE000992; GB:AE000782; NID:g2689315; PIDN:AAB89654.1; PID:g264896

Query Match 4.5%; Score 8; DB 2; Length 72;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 YGLNLNRY 178  
 |||||  
 DB 52 YGLNLNRY 59

RESULT 7  
 E90844  
 probable factor [imported] - Escherichia coli (strain O157:H7, substrain RMD 0509952)  
 C:Species: Escherichia coli  
 C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
 C:Accession: E90844  
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 Gasawara, N.; Yasunaga, T.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene  
 A:Reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: E90844  
 A:Status: preliminary

A:Molecule type: DNA  
 A:Residues: 1-417 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA835148.1; PID:g13361190; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RMD 0509952  
 C:Genetics:  
 A:Gene: ECS1725

Query Match 4.5%; Score 8; DB 2; Length 417;  
 Best Local Similarity 100.0%; Pred. No. 4.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 GLNLNRYF 179  
 |||||  
 DB 256 GLNLNRYF 263

## RESULT 8

D85702  
 hypothetical protein ychP [imported] - Escherichia coli (strain O157:H7, substrain E  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: D85702  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda  
 Nature 403, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: D85702  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-417 <STO>  
 A:Cross-references: GB:AE005174; NID:g12514938; PIDN:AAG56080.1; GSPDB:GN00145; UWG  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: ychP

Query Match 4.5%; Score 8; DB 2; Length 417;  
 Best Local Similarity 100.0%; Pred. No. 4.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 GLNLNRYF 179  
 |||||  
 DB 256 GLNLNRYF 263

## RESULT 9

A64869  
 ychP protein - Escherichia coli (strain K-12)  
 C:Species: Escherichia coli  
 C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
 C:Accession: A64869  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617; PMID:9278503  
 A:Accession: A64869  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-417 <BLAT>

A:Cross-references: GB:AE000220; GB:U00096; NID:g1787467; PIDN:AAC74304.1; PID:g1787  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Genetics:  
 A:Gene: ychP; ychO

Query Match 4.5%; Score 8; DB 2; Length 417;  
 Best Local Similarity 100.0%; Pred. No. 4.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 GLNLNRYF 179  
 |||||  
 DB 256 GLNLNRYF 263

## RESULT 10

F75107 abc transporter, ATP-binding protein PAB1929 - Pyrococcus abyssi (strain Orsay)

C:Species: Pyrococcus abyssi

C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 17-Mar-2000

C:Accession: F75107

R:Anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure

A:Reference number: A75001

A:Accession: F75107

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-633 &lt;KAW&gt;

A:Cross-references: GB:AJ248285; GB:AL096836; NID:g5458067; PIDN:CAB49575.1; PID:e151547

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: PAB1929

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

F:438-623/Domain: ATP-binding cassette homology &lt;ABC&gt;

## Query Match

Best Local Similarity 4.5%; Score 8; DB 2; Length 633;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 KKKHFVET 149

Db 319 KKKHFVET 326

|||||

## RESULT 11

D98068

degenerate transposase (orf1) [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001

C:Accession: D98068

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: D98068

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-66 &lt;KUR&gt;

A:Cross-references: GB:AE007317; PIDN:AAL00377.1; PID:gl5459240; GSPDB:GN00174

C:Genetics:

A:Gene: IS1381-truncation

## Query Match

Best Local Similarity 3.9%; Score 7; DB 2; Length 66;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 SKTHLSA 108

Db 40 SKTHLSA 46

|||||

## RESULT 12

MYTU

myoglobin [validated] - yellowfin tuna

C:Species: Thunnus albacares (yellowfin tuna)

C:Date: 02-Apr-1982 #sequence\_revision 02-Apr-1982 #text\_change 15-Sep-2000

C:Accession: A02526

R:Watts, D.A.; Rice, R.H.; Brown, W.D.

J. Biol. Chem. 255, 10916-10924, 1980

A:Title: The primary structure of myoglobin from yellowfin tuna (Thunnus albacares).

A:Reference number: A02526; MUID:81046956; PMID:7430163

A:Accession: A02526

A:Molecule type: protein

A:Residues: 1-146 &lt;WAT&gt;

R:Birnbaum, G.I.; Evans, S.V.; Przybylska, M.; Rose, D.R.

submitted to the Brookhaven Protein Data Bank, May 1991

A:Reference number: A51306; PDB:1WVT

A:Contents: annotation: X-ray crystallography, 1.74 angstroms, residues 1-146

R:Lattman, E.E.; Nockolds, C.E.; Kretsinger, R.H.; Love, W.E.

J. Mol. Biol. 60, 271-277, 1971

A:Title: Structure of yellow fin tuna metmyoglobin at 6 angstroms resolution.

A:Reference number: A58455; MUID:72014155; PMID:5107328

A:Contents: annotation: X-ray crystallography, 6.0 angstroms

C:Function:

A:Description: binds molecular oxygen for intracellular storage and transport, primer

C:Superfamily: globin; globin homology

C:Keywords: acetylated amino end; chromoprotein; heme; iron; metalloprotein; muscle;

F:1-140/Domain: globin homology &lt;GLB&gt;

F:1/Modified site: acetylated amino end (Ala) #status experimental

F:59/Binding site: oxygen (His) (distal axial ligand) #status experimental

F:88/Binding site: heme iron (His) (proximal axial ligand) #status experimental

## Query Match

Best Local Similarity 3.9%; Score 7; DB 1; Length 146;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 158 VEADYTT 164

Db 13 VEADYTT 19

|||||

## RESULT 13

S00942

hypothetical protein 2 - Halobacterium sp. plasmid pHSB

C:Species: Halobacterium sp.

C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 18-Nov-1994

C:Accession: S00942

R:Kagramanova, V.K.; Derckacheva, N.I.; Mankin, A.S.

Nucleic Acids Res. 16, 4158, 1988

A:Title: The complete nucleotide sequence of the archaeobacterial plasmid pHSB from Hal

A:Reference number: S00941; MUID:88234008; PMID:3287339

A:Accession: S00942

A&gt;Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-168 &lt;KAG&gt;

A:Cross-references: EMBL:X07128

C:Genetics:

A:Genome: plasmid

## Query Match

Best Local Similarity 3.9%; Score 7; DB 2; Length 168;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 75 SRFAIGT 81

Db 26 SRFAIGT 32

|||||

## RESULT 14

AB2849

hypothetical protein Atu2219 [imported] - Agrobacterium tumefaciens (strain C58, Dupo

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

C:Accession: AB2849

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; MCCI

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AB2849

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-180 &lt;KUR&gt;

A:Cross-references: GB:AE008688; PIDN:AAL43208.1; PID:g17740689; GSPDB:GN00186  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: Atu2219  
 A:Map position: circular chromosome

Query Match 3.9%; Score 7; DB 2; Length 180;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 TLRAGLA 96  
 |||||  
 DB 109 TLRAGLA 115

## RESULT 15

F84748  
 probable AP2 domain transcription factor [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: F84748  
 R:Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: F84748  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-218 <STO>  
 A:Cross-references: GB:AE002093; NID:g1707016; PIDN:AAC69127.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g33710  
 A:Map position: 2

Query Match 3.9%; Score 7; DB 2; Length 218;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 LAYDKAA 101  
 |||||  
 DB 107 LAYDKAA 113

## RESULT 16

T02590  
 DNA binding protein EREBP-2 - common tobacco  
 C:Species: Nicotiana tabacum (common tobacco)  
 C:Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 21-Jul-2000  
 C:Accession: T02590  
 R:Ohme-Takagi, M.; Shinshi, H.  
 Plant Cell 7, 173-182, 1995  
 A:Title: Ethylene-inducible DNA binding proteins that interact with an ethylene response  
 A:Reference number: Z14671; MUID:95276459; PMID:7756828  
 A:Accession: T02590  
 A:Status: preliminary; translated from GB/EMBL/DDBJ.  
 A:Molecule type: mRNA  
 A:Residues: 1-233 <OHM>  
 A:Cross-references: EMBL:D38126; NID:g790362; PIDN:BAA07324.1; PID:g1208498  
 A:Experimental source: Strain BY4; tissue-type leaf

Query Match 3.9%; Score 7; DB 2; Length 233;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 LAYDKAA 101  
 |||||  
 DB 137 LAYDKAA 143

## RESULT 17

CRM52

carbonate dehydratase (EC 4.2.1.1) II - mouse  
 N:Alternate names: carbonic anhydrase II  
 C:Species: Mus musculus (house mouse)  
 C:Date: 25-Feb-1985 #sequence\_revision 06-Feb-1995 #text\_change 18-Jun-1999  
 C:Accession: A23900; B23202; A01143; A20539; I51949  
 R:Venta, P.J.; Montgomery, J.C.; Hewett-Emmett, D.; Wiebauer, K.; Tashian, R.E.  
 J. Biol. Chem. 260, 12130-12135, 1985  
 A:Title: Structure and exon to protein domain relationships of the mouse carbonic anhy  
 A:Reference number: A23900; MUID:86008276; PMID:2995362  
 A:Accession: A23900  
 A:Molecule type: DNA  
 A:Residues: 1-260 <VEN>  
 A:Experimental source: strain YBR  
 R:Venta, P.J.; Montgomery, C.; Hewett-Emmett, D.; Tashian, R.E.  
 Biochim. Biophys. Acta 826, 195-201, 1985  
 A:Title: Comparison of the 5' regions of human and mouse carbonic anhydrase II genes  
 A:Reference number: A90655; MUID:86077780; PMID:3000449  
 A:Accession: B23202  
 A:Molecule type: DNA  
 A:Residues: 1-77 <VE2>  
 R:Curtis, P.J.; Withers, E.; Demuth, D.; Watt, R.; Venta, P.J.; Tashian, R.E.  
 Gene 25, 325-332, 1983  
 A:Title: The nucleotide sequence and derived amino acid sequence of cDNA coding for m  
 A:Reference number: A01143; MUID:84109569; PMID:6420240  
 A:Accession: A01143  
 A:Molecule type: mRNA  
 A:Residues: 2-38, 'H', 40-260 <CUR>  
 A:Cross-references: GB:K00811; GB:K00812; GB:M11830; NID:g192333; PIDN:AAA37356.1; PI  
 A:Note: Initiator Met not shown  
 R:Curtis, P.J.  
 J. Biol. Chem. 258, 4459-4463, 1983  
 A:Title: Cloning of mouse carbonic anhydrase mRNA and its induction in mouse erythro  
 A:Reference number: A20539; MUID:83161023; PMID:6187736  
 A:Accession: A20539  
 A:Molecule type: mRNA  
 A:Residues: 155-178;214-240 <CU2>  
 R:Venta, P.J.; Montgomery, J.C.; Wiebauer, K.; Hewett-Emmett, D.; Tashian, R.E.  
 Ann. N. Y. Acad. Sci. 429, 309-323, 1984  
 A:Title: Organization of the mouse and human carbonic anhydrase II genes.  
 A:Reference number: I51949; MUID:84255152; PMID:6331255  
 A:Accession: I51949  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: mRNA  
 A:Residues: 241-260 <RES>  
 A:Cross-references: GB:M25944; NID:g199078; PIDN:AAA39505.1; PID:g199079  
 C:Genetics:  
 A:Gene: Car-2  
 A:Introns: 12/1; 78/1; 117/3; 144/1; 169/3; 221/3  
 C:Superfamily: carbonate dehydratase; carbonic anhydrase homology  
 C:Keywords: carbon-oxygen lyase; hydro-lyase; zinc  
 F:5-259/Domain: carbonic anhydrase homology <CAH>  
 F:94,96,119/Binding site: zinc (HIS) #status predicted  
 Query Match 3.9%; Score 7; DB 1; Length 260;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 YDKAASK 103  
 |||||  
 DB 51 YDKAASK 57

RESULT 18  
 E86354  
 hypothetical protein F16L1.8 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
 C:Accession: E86354  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, L.  
 ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: E86354  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-261 <STO>  
A:Cross-references: GB:AE005172; NID:9454531; PIDN:AAF87854.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 3.9%; Score 7; DB 2; Length 261;  
Best Local Similarity 100.0%; Pred. No. 36; Mismatches 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0;

QY 95 LAYDKAA 101  
|||||  
Db 120 LAYDKAA 126

RESULT 19  
B71877  
hypothetical protein jhp0872 - Helicobacter pylori (strain J99)  
C:Species: Helicobacter pylori  
A:Variety: strain J99  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 08-Oct-1999  
C:Accession: B71877  
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
A:Reference number: A71800; MUID:99120557; PMID:9923682  
A:Accession: B71877  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-274 <ARN>  
A:Cross-references: GB:AE001517; GB:AE001439; NID:94155444; PIDN:AND06452.1; PID:9415544  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: jhp0872

Query Match 3.9%; Score 7; DB 2; Length 274;  
Best Local Similarity 100.0%; Pred. No. 37; Mismatches 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0;

QY 68 KEKYKDN 74  
|||||  
Db 221 KEKYKDN 227

RESULT 20  
F64498  
selenium donor protein - Methanococcus jannaschii  
C:Species: Methanococcus jannaschii  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 17-Mar-2000  
C:Accession: F64498  
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A. Science 273, 1058-1073, 1996  
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
A:Reference number: A64300; MUID:96337999; PMID:8688087  
A:Accession: F64498  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-308 <BUL>  
A:Cross-references: GB:U67599; GB:L77117; NID:91592201; PID:91592202; TIGR:MJ1591; PID:9 C:Genetics:

A:Map position: FOR1563343-1564269  
A:Start codon: TTG  
C:Superfamily: conserved hypothetical protein MJ0640

Query Match 3.9%; Score 7; DB 2; Length 308;  
Best Local Similarity 100.0%; Pred. No. 41; Mismatches 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0;

QY 68 KEKYKDN 74  
|||||  
Db 265 KEKYKDN 271

RESULT 21  
T35643  
glycerol-3-phosphate dehydrogenase - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 11-Jan-2000  
C:Accession: T35643  
R:Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, December 1998  
A:Reference number: Z21585  
A:Accession: T35643  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-336 <MUR>  
A:Cross-references: EMBL:AL034447; PIDN:CAA22402.1; GSPDB:GN00070; SCOEDB:SC7A1.03  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC7A1.03  
C:Superfamily: glycerol-3-phosphate dehydrogenase

Query Match 3.9%; Score 7; DB 2; Length 336;  
Best Local Similarity 100.0%; Pred. No. 44; Mismatches 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0;

QY 90 TLRAGLA 96  
|||||  
Db 86 TLRAGLA 92

RESULT 22  
B30338  
late competence protein comGA - Bacillus subtilis  
N:Alternate names: comG operon protein 1  
C:Species: Bacillus subtilis  
C:Date: 27-Feb-1990 #sequence\_revision 03-Aug-1992 #text\_change 24-Oct-2000  
C:Accession: B30338; B69603  
R:Albano, M.; Breitling, R.; Dubnau, D.A. J. Bacteriol. 171, 5386-5404, 1989  
A:Title: Nucleotide sequence and genetic organization of the Bacillus subtilis comG o  
A:Reference number: A30338; MUID:90008773; PMID:2507524  
A:Accession: B30338  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-356 <ALB>  
A:Cross-references: GB:M29691; GB:M22854; NID:g1119199; PIDN:AAA83367.1; PID:g142706  
A:Note: the authors translated the codon GAT for residue 182 as Glu  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber C.; Bron, S.; Bouilliet, S.; Bruchl, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabet, C.; Ferrari, Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Se akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis

A:Reference number: A69580; MUID:98044033; PMID:9384377  
 A:Accession: B69603  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-356 <KUN>  
 A:Cross-references: GB:299116; GB:AL009126; NID:g2634723; PIDN:CAB14404.1; PID:g2634907  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: comGA  
 C:Superfamily: twitching motility protein pili

Query Match 3.9%; Score 7; DB 2; Length 356;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 GSUTLKL 19  
 |||||  
 DB 77 GSUTLKL 83

## RESULT 23

aminopeptidase homolog [imported] - Halobacterium sp. NRC-1  
 C:Species: Halobacterium sp. NRC-1  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: C84179  
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leitchauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo, J.; Alam, M.; Freitas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; L  
 A:Title: Genome sequence of Halobacterium species NRC-1.  
 A:Reference number: A84160; MUID:20504483; PMID:11016950  
 A:Accession: C84179  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-363 <STO>  
 A:Cross-references: GB:AE004437; NID:g10579833; PIDN:AAG18799.1; GSPDB:GN00138  
 C:Genetics:  
 A:Gene: pepB1

Query Match 3.9%; Score 7; DB 2; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 EADYTK 165  
 |||||  
 DB 70 EADYTK 76

## RESULT 24

hypothetical protein RV3712 - Mycobacterium tuberculosis (strain H37Rv)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C:Accession: A70795  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987; PMID:9634230  
 A:Accession: A70795  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-413 <COL>  
 A:Cross-references: GB:AL022121; GB:AL123456; NID:g3261559; PIDN:CAA18034.1; PID:e136457  
 A:Experimental source: strain H37Rv  
 C:Genetics:  
 A:Gene: RV3712

Query Match 3.9%; Score 7; DB 2; Length 413;

Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 TLRAGLA 96  
 |||||  
 DB 148 TLRAGLA 154

## RESULT 25

probable long-chain fatty acid transport protein fadL [Imported] - Yersinia pestis (s  
 C:Species: Yersinia pestis  
 C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
 C:Accession: AH0334  
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, J.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell  
 Nature 413, 523-527, 2001  
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A:Reference number: AB0001; MUID:21470413; PMID:11586360  
 A:Accession: AH0334  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-423 <KUR>  
 A:Cross-references: GB:AL590842; PIDN:CAC92983.1; PID:g15980722; GSPDB:GN00175  
 C:Genetics:  
 A:Gene: fadL  
 C:Superfamily: long-chain fatty acid transport protein fadL

Query Match 3.9%; Score 7; DB 2; Length 423;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 QWAIHYS 40  
 |||||  
 DB 281 QWAIHYS 287

## RESULT 26

Long-chain fatty acid transport protein precursor STY2623 [Imported] - Salmonella en  
 C:Species: Salmonella enterica subsp. enterica serovar typhi  
 A:Note: this species has also been called Salmonella typhi  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: AE0805  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church, S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, A.; Title: Complete genome sequence of a multiple drug resistant Salmonella enterica s.  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: AE0805  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-437 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD07623.1; PID:g16503614; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY2623  
 C:Superfamily: long-chain fatty acid transport protein fadL

Query Match 3.9%; Score 7; DB 2; Length 437;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 QWAIHYS 40  
 |||||  
 DB 302 QWAIHYS 308

## RESULT 27

Long-chain fatty acid transport protein fadL precursor - Escherichia coli (strain K-

N:Alternate names: FLP protein  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: F65007; A39126; S02829  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: F65007  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-448 <BLAT>  
A:Cross-references: GB:AE000323; GB:U00096; NID:gl1788684; PIDN:AAAC75404.1; PID:gl1788684;  
A:Experimental source: strain K-12, substrain MG1655  
R:Black, P.N.  
J. Bacteriol. 173, 435-442, 1991  
A:Title: Primary sequence of the Escherichia coli fadL gene encoding an outer membrane protein  
A:Reference number: A39126; MUID:91100327; PMID:1987139  
A:Accession: A39126  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-177, 'V', 179-287, 'R', 289-448 <BLA>  
A:Cross-references: GB:M60607; GB:M37714; NID:gl145909; PIDN:AAA64433.1; PID:gl145910  
R:Salz, B.; Ghosh, C.R.; Vu, L.; Nunn, W.D.  
Mol. Microbiol. 2, 363-370, 1988  
A:Title: Nucleotide sequencing and expression of the fadL gene involved in long-chain fatty acid metabolism  
A:Reference number: S02829; MUID:88288050; PMID:2840553  
A:Accession: S02829  
A:Molecule type: DNA  
A:Residues: 66-177, 'R', 179, 'p', 181-287, 'R', 289-448 <SAI>  
A:Cross-references: EMBL:Y00552; NID:gl1371; PIDN:CAA68630.1; PID:gl1372  
A:Note: the authors translated the codon CTG for residue 162 as Glu  
C:Genetics:  
A:Gene: fadL  
C:Superfamily: long-chain fatty acid transport protein fadL  
Query Match 3.9%; Score 7; DB 2; Length 448;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 34 QWAIHYS 40  
|||||||  
Db 313 QWAIHYS 319  
RESULT 28  
long-chain fatty acid transport protein FadL [imported] - Escherichia coli (strain  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: C91032  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: C91032  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-448 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA836650.1; PID:gl13362697; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: ECs3227  
C:Superfamily: long-chain fatty acid transport protein fadL  
Query Match 3.9%; Score 7; DB 2; Length 448;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 34 QWAIHYS 40  
|||||||  
Db 313 QWAIHYS 319  
RESULT 28  
long-chain fatty acid transport protein FadL ECS3227 [imported] - Escherichia coli (strain  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: C91032  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: C91032  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-448 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA836650.1; PID:gl13362697; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: ECs3227  
C:Superfamily: long-chain fatty acid transport protein fadL  
Query Match 3.9%; Score 7; DB 2; Length 448;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 34 QWAIHYS 40  
|||||||  
Db 313 QWAIHYS 319  
RESULT 28  
long-chain fatty acid transport protein FadL - Escherichia coli (strain J99)  
C:Species: Helicobacter pylori  
A:Variety: strain J99  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 08-Oct-1999  
C:Accession: A71859  
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.;  
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.  
Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen  
A:Reference number: A71800; MUID:99120557; PMID:9923682  
A:Accession: A71859  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-511 <ARN>  
A:Cross-references: GB:AE001530; GB:AE001439; NID:gl155602; PIDN:AAD06606.1; PID:gl155602  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: jhp1022  
Query Match 3.9%; Score 7; DB 2; Length 511;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 12 KGSJTLK 18  
|||||||  
Db 283 KGSJTLK 289  
RESULT 31  
F64564  
hypothetical protein HP0358 - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori  
C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
C:Accession: F64564  
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalaf, H.G.; Glodek, A.; McKee, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, Nature 388, 539-547, 1997



A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.  
A:Reference number: A64520; MUID:97394467; PMID:9252185  
A:Accession: F64564  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-511 <TOM>  
A:Cross-references: GB:AE000552; GB:AE000511; NID:g2313451; PIDN:AAD07427.1; PID:g231346

Query Match 3.9%; Score 7; DB 2; Length 511;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 KGSJTLK 18  
|||||  
Db 283 KGSJTLK 289

RESULT 32  
S77572  
oligopeptide transport protein homolog opha - Agrobacterium tumefaciens plasmid pTIR10  
N:Alternate names: oligopeptide permease homolog opha  
C:Species: Agrobacterium tumefaciens  
C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 17-Mar-2000  
C:Accession: S77572  
R:Fuqua, C.; Winans, S.C.  
Mol. Microbiol. 20, 1199-1210, 1996  
A:Title: Localization of OccR-activated and TraR-activated promoters that express two AB  
A:Reference number: S77571; MUID:96405643; PMID:8809772  
A:Accession: S77572  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-519 <FUQ>  
A:Cross-references: EMBL:U48718; NID:g1215729; PIDN:AAC44508.1; PID:g1215731  
A:Experimental source: strain R10  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1996  
C:Genetics:  
A:Gene: opha  
A:Genome: plasmid pTIR10  
C:Superfamily: dipeptide transport protein

Query Match 3.9%; Score 7; DB 2; Length 519;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 90 TLRAGLA 96  
|||||  
Db 89 TLRAGLA 95

RESULT 33  
T08541  
hypothetical protein F27B13.40 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 22-Oct-1999  
C:Accession: T08541  
R:Bevan, M.; Zimmermann, W.; Grueneisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; May  
submitted to the Protein Sequence Database, May 1999  
A:Reference number: Z16442  
A:Accession: T08541  
A:Molecule type: DNA  
A:Residues: 1-526 <BEV>  
A:Cross-references: EMBL:AL050352; GSPDB:GN00062; ATSP:F27B13.40  
A:Experimental source: cultivar Columbia; BAC clone F27B13  
C:Genetics:  
A:Gene: ATSP:F27B13.40  
A:Map position: 4  
A:Introns: 421/3

Query Match 3.9%; Score 7; DB 2; Length 526;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 141 GKXKHEV 147  
|||||  
Db 17 GKXKHEV 23

RESULT 34  
AI0359  
iron(III)-transport system permease sfuB [imported] - *Yersinia pestis* (strain CO92)  
C:Species: *Yersinia pestis*  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
C:Accession: AI0359  
R:Farhnik, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, J.;  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan,  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barral  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AI0359  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-528 <RUR>  
A:Cross-references: GB:AL590842; PIDN:CAC92204.1; PID:g15980916; GSPDB:GN00175  
C:Genetics:  
A:Gene: sfuB  
C:Superfamily: sfuB protein

Query Match 3.9%; Score 7; DB 2; Length 528;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 154 GLLIVEA 160  
|||||  
Db 250 GLLIVEA 256

RESULT 35  
B45808  
B-lymphocyte antigen CD19 precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Sep-1993 #sequence\_revision 02-Aug-1996 #text\_change 02-Aug-2002  
C:Accession: B45808; I49579; I49580  
R:Tedder, T.F.; Isaacs, C.W.  
J. Immunol. 143, 712-717, 1989  
A:Title: Isolation of cDNAs encoding the CD19 antigen of human and mouse B lymphocyt  
A:Reference number: A45808; MUID:89292699; PMID:2472450  
A:Accession: B45808  
A:Molecule type: mRNA  
A:Residues: 302-547 <TED>  
A:Cross-references: GB:M28240; NID:g192827; PIDN:AAA74753.1; PID:g192828  
A:Note: the authors translated the codon GAC for residue 520 as Tyr  
R:Zhou, L.J.; Ord, D.C.; Hughes, A.L.; Tedder, T.F.  
J. Immunol. 147, 1424-1432, 1991  
A:Title: Structure and domain organization of the CD19 antigen of human, mouse, and  
A:Reference number: I48142; MUID:91332460; PMID:1714482  
A:Accession: I49579  
A:Status: translation not shown; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-547 <RES>  
R:Zhou, L.J.; Ord, D.C.; Omori, S.A.; Tedder, T.F.  
Immunogenetics 35, 102-111, 1992  
A:Title: Structure of the genes encoding the CD19 antigen of human and mouse B lymph  
A:Reference number: I49580; MUID:92137946; PMID:1370948  
A:Accession: I49580  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-206, 'A', 208-547 <RE2>  
A:Cross-references: GB:M62553; NID:g192468; PIDN:AAA37390.1; PID:g192470  
C:Genetics:  
A:Gene: CD19  
A:Introns: 30/1; 119/1; 184/1; 277/1; 314/1; 330/2; 360/1; 394/1; 428/1; 451/1; 470/  
C:Superfamily: human B-cell antigen CD19; immunoglobulin homology  
C:Keywords: B-cell; glycoprotein; surface antigen; transmembrane protein

F:31-99/Domain: immunoglobulin homology <IMM1>  
F:190-261/Domain: immunoglobulin homology <IMM2>

Query Match 3.9%; Score 7; DB 2; Length 547;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 QDGSYGE 63  
Db 427 QDGSYGE 433

## RESULT 36

Fiber protein - human adenovirus 31

C:Species: Mastadenovirus h31 (human adenovirus 31)  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 08-Oct-1999  
C:Accession: S39297  
R:Pring-Akerblom, P.; Adrian, T.  
submitted to the EMBL Data Library, November 1993.  
A:Reference number: S39296  
A:Accession: S39297  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-556 <PRI>  
A:Cross-references: EMBL:X76548; NID:g434907; PIDN:CAA54050.1; PID:g434908  
C:Superfamily: adenovirus fiber protein

Query Match 3.9%; Score 7; DB 2; Length 556;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 GSLTLKL 19  
Db 66 GSLTLKL 72

## RESULT 37

A44441

B-cell antigen CD19 precursor - human

N:Alternate names: B-cell-specific glycoprotein CD19; earliest B cell lineage differentiation  
C:Species: Homo sapiens (man)  
C:Date: 30-Sep-1993 #sequence\_revision 02-Aug-1996 #text\_change 02-Aug-2002  
C:Accession: A44441; J10074; A45808; I84731  
R:Kozmik, Z.; Wang, S.; Doerfler, P.; Adams, B.; Busslinger, M.  
Mol. Cell. Biol. 12, 2662-2672, 1992  
A:Title: The promoter of the CD19 gene is a target for the B-cell-specific transcription  
A:Reference number: A44441; MUID:92269839; PMID:1375324  
A:Accession: A44441  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-556 <KOZ>  
A:Cross-references: GB:M84371; NID:g901822; PIDN:AAA69966.1; PID:g901823  
R:Stamenkovic, I.; Seed, B.  
J. Exp. Med. 168, 1205-1210, 1988  
A:Title: CD19, the earliest differentiation antigen of the B cell lineage, bears three e  
A:Reference number: J10074; MUID:89010509; PMID:2459292  
A:Accession: J10074  
A:Molecule type: mRNA  
A:Residues: 1-79, 'I', 81-395, 'RREWAQKKRGRAMRLTVRRTPSSMRTPPTLCRTSSPRMAAATRTLRLMSPWVLRKTE  
A:Cross-references: GB:M21057; NID:g178666; PIDN:AAA35533.1; PID:g178667  
A>Note: the extracellular domain is organized in Ig-like domains  
A>Note: the region beyond residue 395 differs from all other sequence reports in this en  
Epstein-Barr virus BSLF1 protein (see entry PIR1:Q0BE15)  
R:Tedder, T.F.; Isaacs, C.M.  
J. Immunol. 143, 712-717, 1989  
A:Title: Isolation of cDNAs encoding the CD19 antigen of human and mouse B lymphocytes.  
A:Reference number: A45808; MUID:89292699; PMID:2472450  
A:Accession: A45808  
A:Molecule type: mRNA  
A:Residues: 1-513, 'H', 515-556 <TED>  
A:Cross-references: GB:M28170; NID:g862622; PIDN:AAA68490.1; PID:g387662  
R:Zhou, L.J.; Ord, D.C.; Omori, S.A.; Tedder, T.F.

Immunogenetics 35, 102-111, 1992

A:Title: Structure of the genes encoding the CD19 antigen of human and mouse B lympho  
A:Reference number: I49580; MUID:92137946; PMID:1370948  
A:Accession: I84731  
A:Status: translation not shown; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-556 <RES>  
A:Cross-references: GB:M62550; NID:g180032; PID:g180034  
C:Genetics:

A:Gene: GDB:CD19

A:Cross-references: GDB:127605; OMIM:107265

A:Map position: 16p11.2-16p11.2

A:Introns: 30/1; 119/1; 187/1; 279/1; 316/1; 332/2; 362/1; 400/1; 435/1; 458/1; 477/1  
C:Superfamily: human B-cell antigen CD19; immunoglobulin homology  
C:Keywords: B-cell; glycoprotein; surface antigen; transmembrane protein

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-556/Product: B-cell specific glycoprotein CD19 #status predicted <MAT>

F:21-291/Domain: extracellular #status predicted <EXT>

F:193-263/Domain: immunoglobulin homology <IMM>

F:292-319/Domain: transmembrane #status predicted <TMD>

F:310-556/Domain: intracellular #status predicted <INT>

F:86,125,138,181,265/Binding site: carbonylate (Asn) (covalent) #status predicted

Query Match 3.9%; Score 7; DB 2; Length 556;

Best Local Similarity 100.0%; Pred. No. 68;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 QDGSYGE 63  
Db 434 QDGSYGE 440

## RESULT 38

AD1666

DNA ligase homolog lin1870 [imported] - Listeria innocua (strain Clip11262)

C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

C:Accession: AD1666

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec

D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Falhl,

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kraft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;

Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AD1666

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-671 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC97100.1; PID:gl6414371; GSPDB:GN00178

A:Experimental source: strain Clip11262

C:Genetics:

A:Gene: lin1870

C:Superfamily: polydeoxyribonucleotide synthase (NAD+)

Query Match 3.9%; Score 7; DB 2; Length 671;

Best Local Similarity 100.0%; Pred. No. 79;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 92 RAGLAYD 98  
Db 270 RAGLAYD 276

## RESULT 39

T47474

hypothetical protein F18N11.100 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T47474

R:Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unsel, M.; Mewes, H.W.; Lemcke,

submitted to the Protein Sequence Database, February 2000

Search completed: May 12, 2003, 09:49:10  
Job time : 53 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 09:42:05 : Search time 12 seconds  
(without alignments)  
618.689 Million cell updates/sec

Title: US-09-995-493-52

Perfect score: 179

Sequence: 1 QINGVLGPIYGRGSLTKLP.....ADYTKATANLYGLNLNRYF 179

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	7.3	459	1 OM11_HAEIN	P43838 haemophilus
2	13	7.3	459	1 OM12_HAEIN	P10841 haemophilus
3	8	4.5	417	1 YCHO_ECOLI	P39165 escherichia
4	7	3.9	146	1 MYG_GORGI	Q9den9 gobionototh
5	7	3.9	146	1 MYG_THUAL	P02205 thunnus alb
6	7	3.9	146	1 MYG_THUTH	Q9d447 thunnus thy
7	7	3.9	259	1 CAH2_MOUSE	P00920 mus musculu
8	7	3.9	336	1 GPDA_STRCO	Q92850 streptomyc
9	7	3.9	356	1 CMGA_BACSU	P23953 bacillus su
10	7	3.9	415	1 HN3B_ORYLA	O42097 oryzias lat
11	7	3.9	428	1 OM47_PASMU	P80603 pasteurella
12	7	3.9	448	1 FADL_ECOLI	P10384 escherichia
13	7	3.9	466	1 AMID_PSEPU	O69768 pseudomonas
14	7	3.9	547	1 CD19_MOUSE	P25918 mus musculu
15	7	3.9	556	1 CD19_HUMAN	P15391 homo sapien
16	7	3.9	556	1 FIBP_ADE31	P36848 human adeno
17	7	3.9	859	1 RPA2_METJA	Q58446 methanococc
18	7	3.9	865	1 HTRE_ECOLI	P33129 escherichia
19	7	3.9	977	1 A2A1_MOUSE	P17426 m adapter-r
20	7	3.9	1641	1 YC52_XYLEFA	Q9pdx7 xylella fas
21	6	3.4	52	1 ZN32_HUMAN	P17041 homo sapien
22	6	3.4	77	1 RS20_LACLA	Q9ceus5 lactococcus
23	6	3.4	86	1 RS20_MYCLE	P33132 mycobacteri
24	6	3.4	86	1 RS20_MYCTU	P711731 mycobacteri
25	6	3.4	94	1 RL35_PIG	Q29361 sus scrofa
26	6	3.4	97	1 Y004_BHP1	P51703 bacterioph
27	6	3.4	104	1 RK21_PORPU	P51209 porphyra pu
28	6	3.4	120	1 YHFY_ECOLI	P45551 escherichia
29	6	3.4	122	1 RL35_HUMAN	P42766 homo sapien
30	6	3.4	122	1 RL35_RAT	P17078 rattus norv
31	6	3.4	125	1 MSP1_GLORO	P53021 globodera r
32	6	3.4	125	1 MSP2_GLORO	P53022 globodera r
33	6	3.4	125	1 MSP3_GLORO	P53023 globodera r

34	6	3.4	139	1 FXSA_SERMA	P37148 serratia ma
35	6	3.4	146	1 MYG_PSRGE	Q9den1 pseudococ
36	6	3.4	153	1 YXIK_BACSU	P42302 bacillus su
37	6	3.4	180	1 GRPE_STRMU	O06941 streptococ
38	6	3.4	184	1 ITC_ALOMA	P35812 alocasia ma
39	6	3.4	185	1 PTH_BACHD	Q9kgj3 bacillus ha
40	6	3.4	186	1 DYR_HSVS7	P27421 herpesvirus
41	6	3.4	187	1 PABA_ECOLI	P00903 escherichia
42	6	3.4	187	1 PABA_SALTY	P06193 salmonella
43	6	3.4	191	1 SC21_RICPR	Q92804 rickettsia
44	6	3.4	194	1 CLPI_CVAPA	Q36863 cyanophora
45	6	3.4	199	1 SC21_RICCN	Q94jm5 rickettsia
46	6	3.4	202	1 RR41_CVACA	Q94li7 cyanidium c
47	6	3.4	206	1 Y615_SYNY3	P52876 synachocyst
48	6	3.4	210	1 RLA_THETH	P49665 thermus the
49	6	3.4	213	1 DYR_HSVSC	P22573 herpesvirus
50	6	3.4	216	1 ANSP_EIMTE	P15744 eimeria ten
51	6	3.4	224	1 UPP_SYNY3	P72753 synachocyst
52	6	3.4	224	1 RNT_VIBCH	Q9kl97 vibrio chol
53	6	3.4	224	1 UPP_TOBAC	P93394 nicotiana t
54	6	3.4	233	1 ATTB_HVACE	P01512 hyalophora
55	6	3.4	241	1 LEG3_RABIT	P47845 oryctolagus
56	6	3.4	243	1 YC23_CVAPA	P31605 cyanophora
57	6	3.4	244	1 LEG3_CRILO	P47953 cricetus
58	6	3.4	247	1 VATD_BOVIN	P39942 bos taurus
59	6	3.4	247	1 VATD_HUMAN	Q9y5k8 homo sapien
60	6	3.4	247	1 VATD_MOUSE	P57746 mus musculu
61	6	3.4	247	1 VATD_RABIT	Q97755 oryctolagus
62	6	3.4	250	1 H11_DROVI	Q24704 drosophila
63	6	3.4	250	1 H13_DROVI	Q94555 drosophila
64	6	3.4	250	1 VATD_SUBDO	P57747 drosophila
65	6	3.4	259	1 Y348_CHLPN	Q92815 chlamydia p
66	6	3.4	259	1 LEG3_RAT	P08699 rattus norv
67	6	3.4	262	1 TRUA_PYRHO	O58941 pyrococcus
68	6	3.4	263	1 KGRG_ECOLI	P76268 escherichia
69	6	3.4	263	1 LEG3_MOUSE	P16110 mus musculu
70	6	3.4	263	1 PYRF_KLUMA	P41769 kluyveromyc
71	6	3.4	267	1 RM09_HUMAN	Q9byd2 homo sapien
72	6	3.4	267	1 YD24_SCHPO	O13711 schizosacch
73	6	3.4	267	1 TAUD_ECOLI	P37610 escherichia
74	6	3.4	282	1 YQ05_BACAN	Q9rn27 bacillus an
75	6	3.4	286	1 KDSA_FUSNN	Q8re91 fusobacteri
76	6	3.4	286	1 A41_LEIDO	P55905 leishmania
77	6	3.4	288	1 NOLT_RHIFR	P33209 rhizobium f
78	6	3.4	289	1 NOLT_RHIFR	P55714 rhizobium s
79	6	3.4	289	1 TODE_PSEPU	P13453 pseudomonas
80	6	3.4	290	1 VG33_HSVII	Q00118 ictaluriid h
81	6	3.4	290	1 YABG_BACSU	P37548 bacillus su
82	6	3.4	294	1 EFTS_MYCPU	Q98q37 mycoplasma
83	6	3.4	294	1 LEG3_CANFA	P38486 canis famil
84	6	3.4	295	1 RRPP_RABVA	P15198 rabies viru
85	6	3.4	297	1 RRPP_RABVC	P22363 rabies viru
86	6	3.4	297	1 SG84_DROME	Q00725 drosophila
87	6	3.4	297	1 T2B1_BACSU	O68557 bacillus su
88	6	3.4	299	1 HRPS_PSESY	P37931 pseudomonas
89	6	3.4	302	1 OXYR_ERWCA	P71318 erwinia chr
90	6	3.4	302	1 KGRG_ERWCH	P37728 erwinia chr
91	6	3.4	305	1 MEPI_SOYBN	P29136 glycine max
92	6	3.4	305	1 OXYR_ERWCH	Q9x725 erwinia chr
93	6	3.4	305	1 YN45_DEIRA	Q9rry7 deinococcus
94	6	3.4	308	1 VG10_BPB03	Q37891 bacterioph
95	6	3.4	313	1 YK89_ARCFU	O28190 archaeoglob
96	6	3.4	317	1 PIP_SERMA	Q32449 serratia ma
97	6	3.4	319	1 PYRB_MYCTU	P71808 mycobacteri
98	6	3.4	319	1 Y319_MYXAX	P96554 myxococcus
99	6	3.4	321	1 PYRB_MYCLE	Q9cer5 mycobacteri
100	6	3.4	327	1 Y745_HELPJ	Q9z198 helicobacte
101	6	3.4	327	1 Y745_HELPJ	O25441 helicobacte
102	6	3.4	329	1 XYL1_KLUJA	P49378 kluyveromyc
103	6	3.4	335	1 HYPB_RHOCA	P26410 rhodobacter
104	6	3.4	345	1 YF6A_HAEIN	O86241 haemophilus
105	6	3.4	347	1 RTCA_RALSO	Q8y2v6 ralstonia s
106	6	3.4	347	1	

107	6	3.4	353	1	YBGO_ECOLI	P75748	escherichia	180	6	3.4	608	1	HSCA_BUCAP	O51883	buchnera ap
108	6	3.4	358	1	T2H2_HAEPRA	P36433	haemophilus	181	6	3.4	611	1	HSCA_BUCAI	P57660	buchnera ap
109	6	3.4	359	1	RECF_DEIRA	Q9ve0	deinococcus	182	6	3.4	614	1	NCO4_HUMAN	Q13772	homo sapien
110	6	3.4	360	1	ARG1_XENLA	Q91553	xenopus lae	183	6	3.4	621	1	GIDA_HELPJ	Q9zml9	helicobacte
111	6	3.4	360	1	ARG2_XENLA	Q91554	xenopus lae	184	6	3.4	621	1	GIDA_HELPJ	P56138	helicobacte
112	6	3.4	360	1	ARG3_XENLA	Q91555	xenopus lae	185	6	3.4	622	1	AMAI_PLAF8	P50492	plasmodium
113	6	3.4	364	1	DNAB_BORBU	P28616	borrelia bu	186	6	3.4	622	1	AMAI_PLAF8	P50489	plasmodium
114	6	3.4	368	1	P37_MYCCE	Q49410	mycoplasma	187	6	3.4	622	1	AMAI_PLAF8	P22621	plasmodium
115	6	3.4	371	1	DP3B_TREPA	O83048	treponema p	188	6	3.4	622	1	AMAI_PLAF8	P50490	plasmodium
116	6	3.4	371	1	RLPA_ANASP	O8ys24	anabena sp	189	6	3.4	622	1	AMAI_PLAF8	P50491	plasmodium
117	6	3.4	373	1	SWE_HAEIN	P43862	haemophilus	190	6	3.4	644	1	DXS_CHLPN	Q9z6j9	chlamydia p
118	6	3.4	374	1	SBP_CRYJA	P18632	cryptomeria	191	6	3.4	644	1	LONI_METTH	Q36878	methanobact
119	6	3.4	378	1	YH12_MYCTU	Q10400	mycobacteri	192	6	3.4	658	1	CPT2_HUMAN	P23786	homo sapien
120	6	3.4	383	1	GC14_YEAST	P46959	saccharomyc	193	6	3.4	658	1	CPT2_MOUSE	P52825	mus musculus
121	6	3.4	390	1	Y181_ARCFU	Q30056	archaeoglob	194	6	3.4	659	1	AMIA_STRPN	P18791	streptococc
122	6	3.4	392	1	CEMA_NEPOL	Q9tk42	nephroselimi	195	6	3.4	661	1	C180_HUMAN	Q99467	homo sapien
123	6	3.4	393	1	EFTU_CHLVI	P42473	chlorobium	196	6	3.4	662	1	SLST_RABIT	P11770	oryctolagus
124	6	3.4	393	1	LOLC_HAEIN	P44252	haemophilus	197	6	3.4	676	1	LSHR_CALJA	Q02721	callithrix
125	6	3.4	394	1	YIE2_YEAST	P40530	saccharomyc	198	6	3.4	686	1	TGM2_MOUSE	P21981	mus musculus
126	6	3.4	399	1	DNAB_STRCO	P40170	streptomyce	199	6	3.4	689	1	TGM2_MOUSE	P08587	cavia cutle
127	6	3.4	399	1	DXR_CAUCR	Q9a709	caulobacter	200	6	3.4	693	1	TGM3_HUMAN	Q08188	homo sapien
128	6	3.4	401	1	AROC_MYCTU	P95013	mycobacteri	201	6	3.4	693	1	TGM3_MOUSE	Q08189	mus musculus
129	6	3.4	405	1	P152_MOUSE	O70172	mus musculus	202	6	3.4	698	1	ALT_BPT2	Q38424	bacterioph
130	6	3.4	406	1	G6PI_CAMJE	Q9pmd4	campylobact	203	6	3.4	698	1	ALT_BPT6	Q38433	bacterioph
131	6	3.4	406	1	P152_HUMAN	P48426	homo sapien	204	6	3.4	702	1	DDX4_MOUSE	Q61496	mus musculus
132	6	3.4	408	1	P153_HUMAN	P53807	homo sapien	205	6	3.4	702	1	DDX4_MOUSE	Q9hsc3	halobacteri
133	6	3.4	408	1	ODPA_KLULA	Q13366	kluyveromyc	206	6	3.4	713	1	DDX4_RAT	O64060	rattus norv
134	6	3.4	417	1	IRX5_HUMAN	P78411	homo sapien	207	6	3.4	722	1	FLID_TREPA	Q3842	treponema p
135	6	3.4	420	1	GSA_METTH	O26330	methanobact	208	6	3.4	724	1	DDX4_HUMAN	Q9nq10	homo sapien
136	6	3.4	422	1	KICR_MOUSE	P05784	mus musculus	209	6	3.4	724	1	YH76_YEAST	Q05029	saccharomyc
137	6	3.4	432	1	AP2_ARATH	P47927	arabidopsis	210	6	3.4	759	1	MCMS_CAEEL	P21902	caenorhabdi
138	6	3.4	437	1	PEPW_LACDL	P94868	lactobacilli	211	6	3.4	759	1	Y938_MYCTU	P71571	mycobacteri
139	6	3.4	446	1	PPNK_HUMAN	Q95544	homo sapien	212	6	3.4	763	1	PEPX_LACLA	Q9ce01	lactococcus
140	6	3.4	455	1	P2X5_RAT	P51578	rattus norv	213	6	3.4	763	1	PEPX_LACLA	P22346	lactococcus
141	6	3.4	460	1	VATB_METBA	P22663	methanosarc	214	6	3.4	767	1	YNY7_YEAST	P53852	saccharomyc
142	6	3.4	460	1	VATB_METWA	Q60187	methanosarc	215	6	3.4	787	1	YF13_YEAST	P43596	saccharomyc
143	6	3.4	462	1	CBSA_SULAC	Q54088	sulfolobus	216	6	3.4	787	1	YF13_YEAST	P41842	caenorhabdi
144	6	3.4	468	1	PPAI_PICPA	P52291	pichia past	217	6	3.4	793	1	YF13_YEAST	P23377	rattus norv
145	6	3.4	476	1	GLGA_CHLPN	Q9z6v8	chlamydia p	218	6	3.4	796	1	ABAA_EMENI	P20945	emeritella
146	6	3.4	483	1	GABT_CAEEL	Q12127	caenorhabdi	219	6	3.4	806	1	SYFB_MYCCE	P47437	mycoplasma
147	6	3.4	484	1	THII_PSEAE	Q9hu66	pseudomonas	220	6	3.4	811	1	PLSB_VIBCH	Q9kvp8	vibrio chol
148	6	3.4	490	1	CHEI_HUMAN	O00409	homo sapien	221	6	3.4	812	1	FAED_ECOLI	P06970	escherichia
149	6	3.4	490	1	THSA_SULAC	Q9v2t5	sulfolobus	222	6	3.4	833	1	CAFA_YERPE	P28949	yersinia pe
150	6	3.4	491	1	YC46_PORPU	P51189	porphyra pu	223	6	3.4	843	1	RPOC_ODOSI	P49467	odontella s
151	6	3.4	500	1	VJEM_ECOLI	P39282	escherichia	224	6	3.4	861	1	GCR3_YEAST	P34160	saccharomyc
152	6	3.4	506	1	YC46_GUTH	O78439	guillardia	225	6	3.4	873	1	CPHA_SYNY3	P37183	synecocyst
153	6	3.4	510	1	YML4_YEAST	Q03750	saccharomyc	226	6	3.4	903	1	DPOL_BPK11	P38087	bacterioph
154	6	3.4	512	1	BNAL_HUMAN	Q16515	homo sapien	227	6	3.4	906	1	RPOL_BPK11	P18147	bacterioph
155	6	3.4	512	1	BNAL_RAT	Q62962	rattus norv	228	6	3.4	908	1	TB12_NEIMB	Q06987	neisseria m
156	6	3.4	514	1	VS14_TRYBB	P26329	trypanosoma p	229	6	3.4	918	1	CAPP_CORGL	P12880	corynebacte
157	6	3.4	517	1	Y967_TREPA	O83933	treponema p	230	6	3.4	927	1	NED4_HUMAN	P46934	homo sapien
158	6	3.4	529	1	LNT2_RHIME	Q52910	rhizobium m	231	6	3.4	936	1	PMP7_CHLPN	Q9z898	chlamydia p
159	6	3.4	529	1	NOP5_HUMAN	O9y2x3	homo sapien	232	6	3.4	941	1	DNAB_RHOMR	Q30477	rhodothermu
160	6	3.4	530	1	LNT_RHILO	Q98b36	rhizobium l	233	6	3.4	953	1	LKAI_PASHA	P16535	pasteurella
161	6	3.4	531	1	LNT_RHIME	P58377	rhizobium m	234	6	3.4	957	1	NED4_MOUSE	P46935	mus musculus
162	6	3.4	533	1	MLO_HORVU	P93766	hordeum vul	235	6	3.4	960	1	GBR1_MOUSE	Q9w4j8	mus musculus
163	6	3.4	537	1	CYDA_AZOVI	Q09049	azotobacter	236	6	3.4	961	1	GBR1_HUMAN	Q9ub55	homo sapien
164	6	3.4	537	1	DEXB_STREQ	Q59905	streptococc	237	6	3.4	977	1	A2A1_HUMAN	Q95782	h adapter-r
165	6	3.4	551	1	RN27_HUMAN	Q9bzi9	homo sapien	238	6	3.4	991	1	GBR1_RAT	Q9z0u4	rattus norv
166	6	3.4	551	1	RN27_MOUSE	Q99p32	mus musculus	239	6	3.4	992	1	POLS_RUBVM	P08563	rubella vir
167	6	3.4	556	1	TF3B_KLULA	P46070	kluyveromyc	240	6	3.4	1014	1	NANH_CLOSE	P29767	clostridium
168	6	3.4	557	1	G6PI_KLULA	P12341	kluyveromyc	241	6	3.4	1018	1	CONT_HAETN	Q12860	homo sapien
169	6	3.4	563	1	MDLI_PRUSE	P52706	prunus sero	242	6	3.4	1028	1	FDXG_HAETN	P46448	haemophilus
170	6	3.4	568	1	MEND_HAEIN	P44612	h menaquin	243	6	3.4	1046	1	HUUA_HAETN	Q48153	haemophilus
171	6	3.4	575	1	ILVB_LACLA	Q02137	lactococcus	244	6	3.4	1054	1	IF2_STIAM	P55875	stigmattella
172	6	3.4	580	1	SYQ_FALSO	Q8y199	ralstonia s	245	6	3.4	1063	1	POLS_RUBVH	P21480	rubella vir
173	6	3.4	581	1	KAPP_ARATH	P46014	arabidopsis	246	6	3.4	1077	1	HGPA_HAETN	Q9za21	haemophilus
174	6	3.4	585	1	YHD9_YEAST	P38732	saccharomyc	247	6	3.4	1090	1	PULA_KLEPN	P07206	klebsiella
175	6	3.4	587	1	LACP_KLULA	P07921	kluyveromyc	248	6	3.4	1096	1	PULA_KLEAE	P07811	klebsiella
176	6	3.4	596	1	PRIM_CLOAB	P33655	clostridium	249	6	3.4	1103	1	VG37_BPARI	Q9q0b5	bacterioph
177	6	3.4	596	1	TF3B_YEAST	P29056	saccharomyc	250	6	3.4	1118	1	YIPI_YEAST	P40456	saccharomyc
178	6	3.4	598	1	HEXA_ALTSS	P48823	alteromonas	251	6	3.4	1154	1	KDGD_MESAU	Q64398	mesocricetu
179	6	3.4	607	1	YMH1_YEAST	Q04632	saccharomyc	252	6	3.4	1195	1	CHS4_NEUCR	Q01285	neurospora

253	6	3.4	1216	1	RPOB_TROWH	Q93gf2 tropheryma	326	5	2.8	82	1	PSBE_TOBAC	P05168 nicotiana t
254	6	3.4	1250	1	BXE_CLOBO	Q00496 clostridium	327	5	2.8	82	1	PSBE_WHEAT	P05169 triticum ae
255	6	3.4	1250	1	BXE_CLOBO	P30995 clostridium	328	5	2.8	82	1	YWDA_BACSU	P39609 bacillus su
256	6	3.4	1286	1	IRBP_BOVIN	P12661 bos taurus	329	5	2.8	83	1	HPIIS_ARATH	P00262 chromatium
257	6	3.4	1290	1	RPOC_MYCPN	P75271 mycoplasma	330	5	2.8	83	1	PSBE_ARATH	P56779 arabidopsis
258	6	3.4	1296	1	BXG_CLOBO	Q06393 clostridium	331	5	2.8	83	1	PSBE_CHLVU	P56309 chlorella v
259	6	3.4	1338	1	ADO_HUMAN	Q06278 homo sapien	332	5	2.8	83	1	PSBE_GUITH	O78466 guillardia
260	6	3.4	1371	1	UBPU_HUMAN	Q04966 homo sapien	333	5	2.8	83	1	PSBE_ODOSI	P49473 odontella s
261	6	3.4	1375	1	GTFC_STRMU	P13470 streptococc	334	5	2.8	83	1	PSBE_PORPU	P51391 porphyra pu
262	6	3.4	1474	1	SHK2_RAT	Q9qx74 rattus norv	335	5	2.8	84	1	PSBE_CVACA	O9tm20 cyanidium c
263	6	3.4	1476	1	CFTF_MOUSE	P26361 mus musculu	336	5	2.8	85	1	FCEG_BOVIN	O9bdr7 bos taurus
264	6	3.4	1476	1	GTFB_STRMU	Q08987 streptococc	337	5	2.8	85	1	RR16_TOBAC	O06374 nicotiana t
265	6	3.4	1597	1	SOL_DROME	P27398 drosophila	338	5	2.8	86	1	FCEG_CAVPO	Q07249 cavia porce
266	6	3.4	1679	1	YMF9_YEAST	Q04958 saccharomyc	339	5	2.8	86	1	FCEG_HUMAN	P20273 homo sapien
267	6	3.4	2051	1	FAS1_YEAST	P07149 s fatty aci	340	5	2.8	86	1	FCEG_MOUSE	P20491 mus musculu
268	6	3.4	2073	1	FAS1_SCHPO	Q9uu00 s fatty aci	341	5	2.8	86	1	FCEG_PIG	O9xsg6 sus scrofa
269	6	3.4	2133	1	FAB_PIG	P12263 sus scrofa	342	5	2.8	86	1	FCEG_RAT	P20411 rattus norv
270	6	3.4	2210	1	RRPL_EBOSM	Q66802 ebola virus	343	5	2.8	86	1	YABO_BACSU	P37557 bacillus su
271	6	3.4	2319	1	FAB_MOUSE	Q06194 mus musculu	344	5	2.8	86	1	YABO_BACSU	P37557 bacillus su
272	6	3.4	2351	1	FAB_HUMAN	P00451 homo sapien	345	5	2.8	87	1	YABO_BACSU	P37557 bacillus su
273	6	3.4	3298	1	PC16_HUMAN	Q96190 homo sapien	346	5	2.8	88	1	RR16_ECOLI	P76211 escherichia
274	6	3.4	3433	1	UTRO_HUMAN	P46939 homo sapien	347	5	2.8	88	1	RR16_ECOLI	P32087 solanum tub
275	6	3.4	4092	1	DYHC_YEAST	P36022 saccharomyc	348	5	2.8	90	1	HS72_USTMA	P03853 escherichia
276	6	3.4	4351	1	FAT2_RAT	O88277 rattus norv	349	5	2.8	90	1	Y530_BUCAI	P18694 ustilago ma
277	6	3.4	4639	1	DYHC_DROME	P37276 drosophila	350	5	2.8	90	1	Y530_BUCAI	P18694 ustilago ma
278	5	2.8	20	1	CPA7_PAPSP	P80055 papio sp. (	351	5	2.8	90	1	Y144_ARCFU	O28434 archaeoglob
279	5	2.8	22	1	HS71_LEITA	P55938 leishmania	352	5	2.8	91	1	MERP_PSEAE	P04131 pseudomonas
280	5	2.8	24	1	PCL1_PACGO	P82421 pachycondyl	353	5	2.8	91	1	MERP_SHIFL	P04129 shigella fl
281	5	2.8	24	1	PCL2_PACGO	P82422 pachycondyl	354	5	2.8	91	1	RL31_HALNI	O9hmo0 halobacteri
282	5	2.8	36	1	YC12_EUGGR	P31559 euglena gra	355	5	2.8	92	1	Y1PP_KLUJA	P13999 kluyveromyc
283	5	2.8	36	1	PSBI_ARATH	P09370 arabidopsis	356	5	2.8	92	1	IFHE_VIBCH	O9kq44 vibrio chol
284	5	2.8	36	1	PSBI_HORVU	P25876 hordeum vul	357	5	2.8	93	1	S10A_BOVIN	P02639 bos taurus
285	5	2.8	36	1	PSBI_MARPO	P09369 marchantia	358	5	2.8	93	1	S10A_HUMAN	P23237 homo sapien
286	5	2.8	36	1	PSBI_ORISA	P12161 oryza sativ	359	5	2.8	93	1	S10A_MOUSE	P56565 mus musculu
287	5	2.8	36	1	PSBI_PINTH	P41599 pinus thunb	360	5	2.8	94	1	GYRA_STAEP	P54112 staphylococ
288	5	2.8	36	1	PSBI_PSEME	P29796 pseudotsuga	361	5	2.8	94	1	Y950_TREPA	O83916 treponema p
289	5	2.8	39	1	PSBI_SYNY3	Q55354 synechocyst	362	5	2.8	95	1	YCZJ_BACSU	O31846 bacillus su
290	5	2.8	44	1	MFWM_SCHPO	P34069 schizosacch	363	5	2.8	96	1	MPG3_DAGGL	P93124 dactylis gl
291	5	2.8	44	1	PHRA_BACSU	Q00829 bacillus su	364	5	2.8	97	1	FTRV_MAIZE	P80680 zea mays (m
292	5	2.8	47	1	QB54_BACAN	Q9rmx8 bacillus an	365	5	2.8	99	1	CX41_MANSP	O46587 mandrillus
293	5	2.8	49	1	RB12_MOUSE	P35283 mus musculu	366	5	2.8	99	1	CX41_PRECR	O46588 presbytis c
294	5	2.8	51	1	VG12_SPLVR	P15903 spiroplasma	367	5	2.8	99	1	REGN_BPH3	P07243 bacterioph
295	5	2.8	52	1	RK32_GUITH	O78434 guillardia	368	5	2.8	99	1	Y162_STRPY	O99y56 streptococc
296	5	2.8	53	1	RS14_METJA	P54110 methanococc	369	5	2.8	100	1	UCR9_EUGGR	P43266 euglena gra
297	5	2.8	55	1	CX41_CEBAP	O46589 cebus apell	370	5	2.8	101	1	RL36_ENTCP	Q91r88 enteromorph
298	5	2.8	55	1	CX41_SATUS	O46590 saimiri ust	371	5	2.8	102	1	CCAC_AMPCA	P55738 amphidinul
299	5	2.8	56	1	MT_SYNYU	P30565 synechococc	372	5	2.8	102	1	RPOL_AERPE	O9yex1 aeropyrum p
300	5	2.8	56	1	RL32_MYCGE	P47603 mycoplasma	373	5	2.8	103	1	YK85_CAREL	P34351 caenorhabdi
301	5	2.8	59	1	EA57_HUMAN	O43247 homo sapien	374	5	2.8	103	1	NDOA_PSEAE	O51493 pseudomonas
302	5	2.8	64	1	YC33_ODOSI	P49532 odontella s	375	5	2.8	104	1	CYC_RANCA	P23082 pseudomonas
303	5	2.8	65	1	YRP2_SYNPY	Q02191 synechococc	376	5	2.8	104	1	RECA_STRSL	P00024 rana catesb
304	5	2.8	68	1	RL29_STRPN	Q9wv8 streptococc	377	5	2.8	104	1	RL3E_SULAC	P49987 streptococc
305	5	2.8	68	1	RL29_STRPN	O9alw6 streptococc	378	5	2.8	104	1	YBK0_YEAST	P11522 sulfobac
306	5	2.8	70	1	R37A_SULSO	Q97243 sulfobac	379	5	2.8	106	1	RLA3_YEAST	P38168 saccharomyc
307	5	2.8	70	1	R37A_SULSO	Q975h0 sulfobac	380	5	2.8	107	1	COTT_BACSU	P10622 saccharomyc
308	5	2.8	70	1	YL32_HALHA	Q06847 halobacteri	381	5	2.8	107	1	YGLT_ECOLI	P11863 bacillus su
309	5	2.8	71	1	ARRD_RAT	P36577 rattus norv	382	5	2.8	108	1	YGLT_ECOLI	P38521 escherichia
310	5	2.8	73	1	PSBE_CVAPA	P19152 cyanophora	383	5	2.8	109	1	YXIH_BACSU	P42300 bacillus su
311	5	2.8	75	1	YF45_MYCTU	Q10781 mycobacteri	384	5	2.8	110	1	RLA4_YEAST	P02400 saccharomyc
312	5	2.8	77	1	HYEP_PSESP	P80048 pseudomonas	385	5	2.8	112	1	CFP6_MYCTU	O53251 mycobacteri
313	5	2.8	78	1	KEAI_ECOLI	Q52278 escherichia	386	5	2.8	113	1	CYC_CRIFA	P00078 crithidia f
314	5	2.8	78	1	YO09_BPL2	P42544 bacterioph	387	5	2.8	113	1	GVKT_HALNI	P24375 halobacteri
315	5	2.8	80	1	PSBE_MESVI	Q9muq0 mesostigma	388	5	2.8	113	1	RL19_MYCLE	O33020 mycobacteri
316	5	2.8	80	1	PSBE_SYNY3	P09190 synechocyst	389	5	2.8	113	1	RL19_MYCLE	Q10792 mycobacteri
317	5	2.8	82	1	DNAL_STRAL	P96457 streptocyst	390	5	2.8	113	1	RLA1_DICDI	P22684 dictyosteli
318	5	2.8	82	1	PSBE_CHLRE	P48268 chlamydomon	391	5	2.8	113	1	V186_FOWPV	Q9j548 fowlpox vir
319	5	2.8	82	1	PSBE_MARPO	P06851 marchantia	392	5	2.8	113	1	Y655_FACHD	Q9z9w1 bacillus ha
320	5	2.8	82	1	PSBE_MESCR	P36442 mesembryant	393	5	2.8	114	1	RK20_CHLVU	P56352 chlorella v
321	5	2.8	82	1	PSBE_OENBE	P05170 oenothera b	394	5	2.8	114	1	RLA2_EIMTE	Q967y9 elmeria ten
322	5	2.8	82	1	PSBE_OENHO	Q9mtk5 oenothera h	395	5	2.8	115	1	CMGE_BACSU	P25957 bacillus su
323	5	2.8	82	1	PSBE_PEN	P13554 pisum sativ	396	5	2.8	115	1	PTHY_RAT	P04089 rattus norv
324	5	2.8	82	1	PSBE_PINTH	P41615 pinus thunb	397	5	2.8	115	1	RL19_BACSU	O31742 bacillus su
325	5	2.8	82	1	PSBE_SPIOL	P09197 spinacia ol	398	5	2.8	116	1	NU3M_MACRO	P92666 macropus ro
326	5	2.8	82	1	PSBE_SPIOL	P09197 spinacia ol	399	5	2.8	116	1	RL19_MACRO	P30529 bacillus st

399	5	2.8	116	1	RL19_STRCO	O69883 streptomyc	472	5	2.8	131	1	YQ05_PSEAE	Q910n3 pseudomonas
400	5	2.8	116	1	RNT_AZOVI	Q925x0 azotobacter	473	5	2.8	132	1	CRL_ECOLI	P245n1 escherichia
401	5	2.8	117	1	GVPK_HALME	Q02236 halobacteri	474	5	2.8	132	1	NIFW_FRAAL	P46039 frankia aln
402	5	2.8	117	1	R18E_ARCFU	O29138 archaeoglob	475	5	2.8	134	1	Y605_METJA	Q58022 methanococc
403	5	2.8	118	1	NLT6_AMBAR	O04004 ambrosia ar	476	5	2.8	135	1	VLYS_BPPH1	Q92xd8 bacterioph
404	5	2.8	118	1	RLI7_AQAE	O66482 aquiflex aeo	477	5	2.8	136	1	BLS_STRSJ	P19997 streptovir
405	5	2.8	118	1	RLI7_THETH	Q929m5 thermus the	478	5	2.8	136	1	RR16_CHLSP	P05727 chlamydomon
406	5	2.8	119	1	ACPS_HELPJ	Q92136 helicobacte	479	5	2.8	136	1	RR16_CHLSP	P05727 chlamydomon
407	5	2.8	119	1	GVKS_HELPJ	O25488 helicobacte	480	5	2.8	136	1	RL27_CAREL	P19194 caenorhabdi
408	5	2.8	119	1	WNT4_HALN1	P33963 halobacteri	481	5	2.8	137	1	FX41_PERPO	O46586 perodicticu
409	5	2.8	119	1	RL14_PARTE	P15767 paramecium	482	5	2.8	137	1	PL15A_RABIT	P26202 oryctolagus
410	5	2.8	119	1	SSNA_HUMAN	O43805 homo sapien	483	5	2.8	137	1	PL15B_RABIT	P26203 oryctolagus
411	5	2.8	119	1	WNT4_CHECA	P28108 chen caerul	484	5	2.8	138	1	NRDI_MYCLE	Q9cbp9 mycobacteri
412	5	2.8	119	1	WNT4_EUMSK	P28116 eumeces ski	485	5	2.8	138	1	RBS_PORPU	P10272 porphyra pu
413	5	2.8	119	1	WNT4_MELGA	P28127 meleagris g	486	5	2.8	138	1	RL29_MYCCA	P11242 mycoplasma
414	5	2.8	119	1	WNT4_PLEJO	P28135 plethodon j	487	5	2.8	138	1	RS8_THETH	P24319 thermus the
415	5	2.8	119	1	WNT4_SCEOC	P28143 sceloporos	488	5	2.8	138	1	YBE9_YEAST	P38191 saccharomyc
416	5	2.8	120	1	NLTP_DAUCA	P27631 daucus caro	489	5	2.8	139	1	RBS_OLILU	P14961 olisthodisc
417	5	2.8	120	1	RL7A_HALN1	Q9hgh8 halobacteri	490	5	2.8	139	1	YOGF_HAEIN	P43981 haemophilus
418	5	2.8	120	1	RS15_NAEGR	P51429 naegleria g	491	5	2.8	140	1	CSF2_CAVPO	Q50481 cavia porce
419	5	2.8	120	1	VATE_SCHPO	O43046 schizosacch	492	5	2.8	141	1	GLMB_HUMAN	P17774 homo sapien
420	5	2.8	120	1	YE07_SCHPO	O13804 schizosacch	493	5	2.8	141	1	GLMB_RAT	O63228 rattus norv
421	5	2.8	121	1	RS24_PYRAE	Q82v65 pyrobaculum	494	5	2.8	141	1	HBA1_EQUBU	Q9xsn3 equus burch
422	5	2.8	121	1	SHH_AMECH	P79682 amblypharyn	495	5	2.8	141	1	HBA2_EQUBU	Q9tva3 equus burch
423	5	2.8	121	1	SHH_CARAU	P79691 carassius a	496	5	2.8	141	1	HBA7_HORSE	P06714 equus cabal
424	5	2.8	121	1	SHH_DANAA	O13234 danio aff.	497	5	2.8	141	1	HBA_AILFU	P18969 allurus ful
425	5	2.8	121	1	SHH_DANAE	O13234 danio aff.	498	5	2.8	141	1	HBA_BRATR	P14525 bradyus tr
426	5	2.8	121	1	SHH_DANAT	O13238 danio aff.	499	5	2.8	141	1	HBA_CERAE	P01926 cercopithe
427	5	2.8	121	1	SHH_DANR	O13245 danio frank	500	5	2.8	141	1	HBA_CERTO	P01933 cercocebus
428	5	2.8	121	1	SHH_DANKE	P79709 danio kerri	501	5	2.8	141	1	HBA_COLBA	P01980 colobus bad
429	5	2.8	121	1	SHH_DANPU	P79717 danio pulch	502	5	2.8	141	1	HBA_DASNO	P01984 dasypus nov
430	5	2.8	121	1	SHH_DEVDE	O13241 devario dev	503	5	2.8	141	1	HBA_ECHTE	P24291 echinops te
431	5	2.8	121	1	SHH_DEVNA	O13247 devario mal	504	5	2.8	141	1	HBA_EQUAS	P01959 equus asinu
432	5	2.8	121	1	SHH_DEVPA	O13250 devario pat	505	5	2.8	141	1	HBA_EQUHE	P01961 equus hemio
433	5	2.8	121	1	SHH_PUNCO	P79838 puntius con	506	5	2.8	141	1	HBA_EQUPR	Q9xse9 equus cabal
434	5	2.8	121	1	SHH_PUNTE	P79850 puntius tet	507	5	2.8	141	1	HBA_EQUZE	P01960 equus cabal
435	5	2.8	121	1	SHH_RASEL	P79858 rasbora ele	508	5	2.8	141	1	HBA_HORSE	P01958 equus cabal
436	5	2.8	121	1	SHH_RASER	P79864 rasbora het	509	5	2.8	141	1	HBA_MACAS	P21766 macaca assa
437	5	2.8	121	1	SHH_RASPA	P79869 rasbora pav	510	5	2.8	141	1	HBA_MACFA	P21767 macaca fasc
438	5	2.8	121	1	SHH_TANAL	P79915 tanichthys	511	5	2.8	141	1	HBA_MACMU	P01925 macaca mula
439	5	2.8	121	1	Y38_BPT7	P03797 bacterioph	512	5	2.8	141	1	HBA_MACNE	P19002 macaca neme
440	5	2.8	122	1	H2B_PLADU	P19374 platynereis	513	5	2.8	141	1	HBA_MACSI	P21768 macaca sinl
441	5	2.8	122	1	H2B_SIPNU	P30757 sipunculus	514	5	2.8	141	1	HBA_MACSP	P07402 macaca spec
442	5	2.8	122	1	H2B_URECA	P27326 urechis cau	515	5	2.8	141	1	HBA_MANSF	P08258 mandrillus
443	5	2.8	122	1	MPT2_PHLPR	P43214 phleum prat	516	5	2.8	141	1	HBA_MARFO	P17689 martes foli
444	5	2.8	122	1	NTF2_ORYSA	Q9xj54 oryza sativ	517	5	2.8	141	1	HBA_PAGLA	P19645 paguma larv
445	5	2.8	122	1	RL14_MYCGE	P47407 mycoplasma	518	5	2.8	141	1	HBA_PAPCY	P01931 papio cynoc
446	5	2.8	122	1	TVB5_MOUSE	P04213 mus musculu	519	5	2.8	141	1	HBA_PROLO	P18977 procyon lot
447	5	2.8	122	1	YEBE_ECOLI	P33219 escherichia	520	5	2.8	141	1	HBA_THEGE	P01932 theropithe
448	5	2.8	124	1	CC26_YEAST	P14724 saccharomyc	521	5	2.8	141	1	HIS3_METKA	P38836 methanopyru
449	5	2.8	124	1	CHEY_HELPY	P71403 helicobacte	522	5	2.8	141	1	PRO1_STRPU	P32006 strongyloce
450	5	2.8	124	1	CX41_SATSC	O46582 saimiri sci	523	5	2.8	141	1	Y427_MYCPN	P75170 mycoplasma
451	5	2.8	124	1	GC5H_THCAC	Q9h1h3 thermoplas	524	5	2.8	141	1	YJEB_ECOLI	P21498 escherichia
452	5	2.8	126	1	NIFW_FRASE	Q925y0 frankia sp	525	5	2.8	142	1	CU07_MOUSE	P58500 mus musculu
453	5	2.8	126	1	PFDA_CAREL	O17435 caenorhabdi	526	5	2.8	142	1	GLMG_HUMAN	O60234 homo sapien
454	5	2.8	126	1	PLAS_SYNY3	P21697 synecocyst	527	5	2.8	142	1	MMSS_MYCTU	O53785 mycobacteri
455	5	2.8	126	1	Y854_PYRHO	O58584 pyrococcus	528	5	2.8	142	1	RL13_BUCAL	P57471 buchnera ap
456	5	2.8	126	1	YP7_AGRTO	P03867 agrobacteri	529	5	2.8	142	1	RL25_KLULA	P48045 kluyveromyc
457	5	2.8	128	1	CD59_AOTTR	P51447 actus trivi	530	5	2.8	142	1	YK24_YEAST	P53143 saccharomyc
458	5	2.8	128	1	CD59_CERAE	Q28216 cercopithe	531	5	2.8	143	1	FRDA_ARATH	Q92r07 arabidopsis
459	5	2.8	128	1	CD59_HUMAN	P13987 h cd59 glyc	532	5	2.8	143	1	PSGD_HUMAN	Q9uq72 homo sapien
460	5	2.8	128	1	DAG1_PIG	Q29243 sus.scrofa	533	5	2.8	143	1	PTMA_MYCPN	P75145 mycoplasma
461	5	2.8	128	1	LY6D_HUMAN	Q14210 homo sapien	534	5	2.8	143	1	RK2_SOYBN	P18663 glycine max
462	5	2.8	129	1	RL3A_AEDAL	P45842 aedes albop	535	5	2.8	144	1	CX41_AOTAZ	O46584 actus azara
463	5	2.8	129	1	RS1A_ARATH	P42798 arabidopsi	536	5	2.8	144	1	CX41_GORGO	O46578 gorilla gor
464	5	2.8	129	1	RS1A_BRANA	Q00332 brassica na	537	5	2.8	144	1	CX41_PANTR	O46577 pan troglod
465	5	2.8	130	1	EXR3_ARATH	Q92v52 arabidopsi	538	5	2.8	144	1	CX41_PITPI	O46585 pithecia pi
466	5	2.8	130	1	YERA_YEREN	P31490 yersinia en	539	5	2.8	144	1	CX41_PONPY	O46579 pongo pygma
467	5	2.8	130	1	YERA_YERPE	P31491 yersinia pe	540	5	2.8	144	1	CX41_THEGE	O46581 theropithe
468	5	2.8	131	1	ANFB_PIG	P07634 sus.scrofa	541	5	2.8	144	1	FUCU_HAEIN	P44778 haemophilus
469	5	2.8	131	1	CRB2_SATSC	P47777 saimiri sci	542	5	2.8	144	1	HMGL_PHANI	P40619 pharbitis n
470	5	2.8	131	1	CRB2_BACSU	O07591 bacillus su	543	5	2.8	144	1	MERR_PSEAE	P06688 pseudomonas
471	5	2.8	131	1	RS4_MYCBO	P45811 mycobacteri	544	5	2.8	144	1	PSGB_HUMAN	Q9uq74 homo sapien



545	1	YHP5_YEAST	P38808	saccharomyc	618	5	2.8	164	1	PTPA_STRCO	P53433	streptomyce
546	1	DUT_PSEAE	Q9hu44	pseudomonas	619	5	2.8	164	1	TKD1_BOVIN	Q28201	bos taurus
547	1	DUT_CHLMU	Q9pka2	chlamydia m	620	5	2.8	164	1	VATL_PLECA	Q43362	pleurochrys
548	1	DUT_CHLTR	O84294	chlamydia t	621	5	2.8	164	1	Y338_STRPY	Q9a1d3	streptococ
549	1	PA2M_CAVPO	P47711	cavia porce	622	5	2.8	165	1	DEST_CHICK	P18359	gallus gall
550	1	CYNS_ANASP	P58703	anabaena sp	623	5	2.8	165	1	DSBB_VIBAL	Q56378	vibrio algi
551	1	RS19_ORISA	P40978	oryza sativ	624	5	2.8	165	1	MLRV_MOUSE	P51567	mus musculu
552	1	VPGL_HAEIN	O05072	haemophilus	625	5	2.8	165	1	MLRV_MOUSE	P08733	rattus norv
553	1	XKDJ_BACSU	P54330	bacillus su	626	5	2.8	165	1	YB87_YEAST	P38362	saccharomyc
554	1	YAAR_BACSU	P37539	bacillus su	627	5	2.8	165	1	YG06_YEAST	P53095	saccharomyc
555	1	YB99_METJA	Q58599	methanococ	628	5	2.8	165	1	YRHL_RHOSH	Q53229	rhodobacter
556	1	ILA_RAT	P20096	rattus norv	629	5	2.8	165	1	YXEL_BACSU	P54951	baecillus su
557	1	YSMA_BACSU	P11469	bacillus su	630	5	2.8	166	1	YXEL_BACSU	P29779	onchocerca
558	1	CALM_EUGGR	P11118	euglena gra	631	5	2.8	166	1	OV22_ONCVO	Q9rbw4	helicobacte
559	1	CALM_TRYBB	P04465	trypanosoma	632	5	2.8	166	1	TPX_HELPJ	Q9zke7	delinococcus
560	1	CALM_TRYCR	P18061	trypanosoma	633	5	2.8	166	1	TPX_HELPJ	Q9zke7	delinococcus
561	1	RM27_MOUSE	Q99p92	mus musculu	634	5	2.8	167	1	ELBS_ADE40	Q25151	helicobacte
562	1	YEMA_YEAST	P40030	saccharomyc	635	5	2.8	167	1	YSX5_CAEEL	P10343	human adeno
563	1	YISS_YEAST	P40565	saccharomyc	636	5	2.8	168	1	PTP_NPVAC	Q10024	caenorhabdi
564	1	HMGL_VICFA	P40620	vicia faba	637	5	2.8	168	1	SP22_ORISA	P24656	autographa
565	1	MAFF_CHICK	Q90595	gallus gall	638	5	2.8	168	1	TRIC_CHICK	Q919b4	oryza sativ
566	1	NORC_PARHA	O50651	paracoccus	639	5	2.8	168	1	YCD3_YEAST	P27673	gallus gall
567	1	CRPD_CHLTR	O84449	chlamydia t	640	5	2.8	169	1	CX41_HUMAN	P25566	saccharomyc
568	1	CRPL_CHLTR	P18587	chlamydia t	641	5	2.8	169	1	RECA_STRMU	P13073	homo sapien
569	1	MOAE_HAEIN	P45308	haemophilus	642	5	2.8	169	1	RL10_RICCN	P27624	streptococ
570	1	RL9_STRPN	Q97n63	streptococ	643	5	2.8	169	1	RL10_RICCN	Q92388	rickettsia
571	1	SSRP_THEMEA	P56944	thermotoga	644	5	2.8	169	1	SNX0_HUMAN	Q9z622	rickettsia
572	1	UL96_HSVSA	Q01024	herpesvirus	645	5	2.8	170	1	ELBS_ADE41	Q9y343	homo sapien
573	1	YF13_MYCPN	P75273	mycoplasma	646	5	2.8	170	1	OCRA_STRLI	P10544	human adeno
574	1	CRPE_CHLTR	P26757	chlamydia t	647	5	2.8	171	1	AK15_HUMAN	Q9z6b7	streptomyce
575	1	ARPD_CRAPL	P22441	craterostig	648	5	2.8	171	1	ATPF_SYN6	Q9nph9	homo sapien
576	1	AROO_SYNY3	P73367	synechocyst	649	5	2.8	171	1	OMPX_ECOLI	P08447	synechococ
577	1	CRPA_CHLMU	Q9p1v1	chlamydia m	650	5	2.8	172	1	OLE2_ORISA	P36546	escherichia
578	1	CRPC_CHLTR	P26756	chlamydia t	651	5	2.8	172	1	PHB1_PREDI	Q40646	oryza sativ
579	1	SP6_MOUSE	Q9esx2	mus musculu	652	5	2.8	172	1	RL10_CAUCR	P07119	fremyella d
580	1	MLR_PATSP	P02613	patinopecte	653	5	2.8	172	1	RUNC_ECOLI	P58060	caulobacter
581	1	TPC_TACTR	P15159	tachypleus	654	5	2.8	172	1	TCYP_DROME	P24339	escherichia
582	1	YGH4_ECOLI	P33354	escherichia	655	5	2.8	172	1	Y096_NPVOP	Q9vgs2	drosophila
583	1	YEG4_YEAST	P53342	saccharomyc	656	5	2.8	173	1	RUNC_SALTY	Q83951	orgyia pseu
584	1	SSRP_LISMO	Q928j0	listeria mo	657	5	2.8	173	1	RUNC_YERPE	Q8xgf7	salmonella
585	1	SSRP_STAAM	Q99vk0	staphylococ	658	5	2.8	173	1	YKNI_YEAST	Q8zeu7	versinia pe
586	1	YHCH_ECOLI	P45424	escherichia	659	5	2.8	174	1	ENV_AVISU	P36067	saccharomyc
587	1	RL22_PYRAB	Q9vju0	pyrococcus	660	5	2.8	174	1	NADD_HELPJ	P03398	avian sarco
588	1	RR7_EPIVI	O59423	pyrococcus	661	5	2.8	174	1	NADD_HELPJ	Q923p8	helicobacte
589	1	RS7_UREPA	P30057	epifagus vi	662	5	2.8	175	1	IF3_CLOAB	O25895	helicobacte
590	1	Y150_TREPA	Q9ppw6	ureaplasma	663	5	2.8	176	1	RET1_ONCMY	Q97gk5	clostridium
591	1	MLRB_PATYE	O83185	treponema p	664	5	2.8	176	1	RET2_ONCMY	P24774	oncorhynch
592	1	MLRC_PATYE	P04112	patinopecte	665	5	2.8	177	1	CYCL_PARDE	P24775	oncorhynch
593	1	MLR_AEQIR	P05944	patinopecte	666	5	2.8	177	1	DSBB_HAEIN	P29899	paracoccus
594	1	MLR_CHLNI	P13543	aequipecten	667	5	2.8	177	1	YDAB_MYCBO	P44707	haemophilus
595	1	NUSB_MYCTU	P05963	chlamys nip	668	5	2.8	178	1	ESM5_DROME	P46830	mycobacteri
596	1	Y498_METJA	P95020	mycobacteri	669	5	2.8	178	1	HPRT_ECOLI	P13096	drosophila
597	1	BFRB_NEIGO	Q57921	methanococ	670	5	2.8	178	1	MUPL_MOUSE	P36766	escherichia
598	1	BFRB_NEIMA	P77914	neisseria g	671	5	2.8	179	1	PRU_MYXXA	Q9d267	mus musculu
599	1	PSAA_YERPE	P56999	neisseria m	672	5	2.8	179	1	RL5_RICPR	P27755	myxococcus
600	1	PSAA_YERPS	P31522	versinia pe	673	5	2.8	179	1	RL6_STRCO	Q9zcr7	rickettsia
601	1	YRS3_CAEEL	Q56982	versinia ps	674	5	2.8	180	1	ATPD_ENTHR	P46786	streptomyce
602	1	AL12_APIGR	O09347	caenorhabdi	675	5	2.8	180	1	RL5_MYCCA	P26680	enterococcu
603	1	PRZA_MYXXA	P92918	apium grave	676	5	2.8	180	1	TRIC_TRIVU	P10136	mycoplasma
604	1	IFSA_DIACA	P43498	myxococcus	677	5	2.8	181	1	Y4WC_RHISN	Q29147	trichosurus
605	1	RJ5B_MYCTU	Q9axg7	dianthus ca	678	5	2.8	181	1	APT_STRCO	P55685	rhizobium s
606	1	YJ59_AQUAE	P71685	mycobacteri	679	5	2.8	182	1	ATPD_BACFI	P52561	streptomyce
607	1	BFR_BRUME	O67771	aquifex aeo	680	5	2.8	182	1	C560_CAEEL	P22779	baecillus fi
608	1	RNPA_HELPY	P49944	brucella me	681	5	2.8	182	1	GVH1_HALN1	P41956	caenorhabdi
609	1	VAL2_CAEEL	P55997	helicobacte	682	5	2.8	182	1	PDAD_THEAC	P24372	halobacteri
610	1	VATL_ASCSU	P34546	caenorhabdi	683	5	2.8	182	1	YE03_HAEIN	Q9hk30	thermoplas
611	1	Y255_AQUAE	Q17046	ascaris suu	684	5	2.8	183	1	AACL_DICDI	P44178	haemophilus
612	1	MCRW_METTH	O66616	aquifex aeo	685	5	2.8	183	1	CITX_ECOLI	P14195	dictyostelli
613	1	PHEA_ANASP	Q50485	methanobact	686	5	2.8	183	1	RL6_CHLUPN	P77563	escherichia
614	1	SKP1_HUMAN	P35796	anabaena sp	687	5	2.8	183	1	RL6_CHLUPN	Q9p1m8	chlamydia m
615	1	SODM_CORDI	P34991	homo sapien	688	5	2.8	183	1	RL6_CHLTR	Q9z7s1	chlamydia p
616	1	PUR6_PSEAE	P42821	corynebacte	689	5	2.8	183	1	YG1L_ECOLI	P25056	chlamydia t
617	1		P72157	pseudomonas	690	5	2.8	183	1		Q11153	mycobacteri
											P39834	escherichia

691	1	ATPE_SPIOL	184	2.8	5	764	1	TORD_ECOLI	199	2.8	5	P06453 spinacia ol
692	1	C560_CABER	184	2.8	5	765	1	Y599_METJA	199	2.8	5	P41955 caenorhabdi
693	1	YDJA_HAEIN	184	2.8	5	766	1	COEL_CORGL	200	2.8	5	P45244 haemophilus
694	1	KADA_METTH	185	2.8	5	767	1	YCLP_XANCP	200	2.8	5	O26135 methanobact
695	1	NUSG_HAEIN	185	2.8	5	768	1	LASI_PSEAE	201	2.8	5	P43916 haemophilus
696	1	YPI3_VIBAL	185	2.8	5	769	1	MINC_AQUAE	201	2.8	5	P52059 vibrio algi
697	1	YPRE_PYRHO	186	2.8	5	770	1	RECR_AGR5	201	2.8	5	O58955 pyrococcus
698	1	YP89_CABEL	186	2.8	5	771	1	RECR_BRUME	201	2.8	5	O09227 caenorhabdi
699	1	LOE3_MAIZE	187	2.8	5	772	1	RL15_SCHPO	201	2.8	5	P21641 zea mays (m
700	1	SODC_HAEIN	187	2.8	5	773	1	RS4_MYCLE	201	2.8	5	P25841 haemophilus
701	1	SODC_HAEPA	187	2.8	5	774	1	RS4_MYCTU	201	2.8	5	P25842 haemophilus
702	1	UL43_HCWA	187	2.8	5	775	1	SODF_METVJ	202	2.8	5	P16781 human cytom
703	1	ADML_PIG	188	2.8	5	776	1	T4S1_HUMAN	202	2.8	5	P53366 sus scrofa
704	1	EPF_UREPA	188	2.8	5	777	1	UMPK_ARATH	202	2.8	5	O9pqj3 ureaplasma
705	1	SLP_ECOLI	188	2.8	5	778	1	Y063_HUMAN	202	2.8	5	P37194 escherichia
706	1	TFT1_PIG	188	2.8	5	779	1	AG35_VACCV	203	2.8	5	O97557 sus scrofa
707	1	TRAF_RHSN	188	2.8	5	780	1	FLA1_ARCFU	203	2.8	5	P55417 rhizobium s
708	1	YLAI_LACAC	188	2.8	5	781	1	FLA1_ARCFU	203	2.8	5	P29470 lactobacill
709	1	ICYA_MANSE	189	2.8	5	782	1	SODM_PSEPU	203	2.8	5	P00305 manduca sex
710	1	OXYR_MYCXE	189	2.8	5	783	1	SODM_XANCP	203	2.8	5	O87883 mycobacteri
711	1	RPOE_HAEIN	189	2.8	5	784	1	YGIN_HAEIN	203	2.8	5	P44790 haemophilus
712	1	GCHI_CAMJE	190	2.8	5	785	1	CRB2_MOUSE	204	2.8	5	P51594 campylobact
713	1	H5_METJA	190	2.8	5	786	1	GIDB_COXBU	204	2.8	5	P54040 mycobaccoc
714	1	HPPK_MYCLE	191	2.8	5	787	1	NADD_TREPA	204	2.8	5	O69528 mycobacteri
715	1	RPOE_ECOLI	191	2.8	5	788	1	CYSR_SINY3	205	2.8	5	P34086 escherichia
716	1	RPOE_SALTY	191	2.8	5	789	1	IL6_ORCOR	205	2.8	5	P37401 salmonella
717	1	SSUE_ECOLI	191	2.8	5	790	1	TNFB_HUMAN	205	2.8	5	P80644 escherichia
718	1	APL2_ARATH	192	2.8	5	791	1	YPT7_SCHPO	205	2.8	5	O42563 arabidopsis
719	1	FLHC_ECOLI	192	2.8	5	792	1	EFA4_MOUSE	206	2.8	5	P11165 escherichia
720	1	FLHC_SALTY	192	2.8	5	793	1	H1T_HUMAN	206	2.8	5	O52222 salmonella
721	1	HIS7_STAAM	192	2.8	5	794	1	ICYE_MANSE	206	2.8	5	O99445 staphylococ
722	1	KPY1_SPICI	192	2.8	5	795	1	RR4_ASTLO	206	2.8	5	P19680 spiroplasma
723	1	RM11_HUMAN	192	2.8	5	796	1	COAE_DEIRA	207	2.8	5	O9y3b7 homo sapien
724	1	YOHJ_ECOLI	192	2.8	5	797	1	H1T_MACMU	207	2.8	5	P33366 escherichia
725	1	FLHC_ERWCA	193	2.8	5	798	1	IL6_CANFA	207	2.8	5	O9x601 erwinia car
726	1	FLHC_PROMI	193	2.8	5	799	1	IL6_MARMO	207	2.8	5	O34202 proteus mir
727	1	FLHC_YEREN	193	2.8	5	800	1	IM22_YEAST	207	2.8	5	O86047 yersinia en
728	1	H10_HUMAN	193	2.8	5	801	1	PSMB_PYRHO	207	2.8	5	P07305 homo sapien
729	1	H10_MOUSE	193	2.8	5	802	1	RL4_BACST	207	2.8	5	P10922 mus musculu
730	1	H10_RAT	193	2.8	5	803	1	TRIC_COTJA	207	2.8	5	P43278 rattus norv
731	1	RHOA_CANFA	193	2.8	5	804	1	UPP_MYCTU	207	2.8	5	P24406 canis famil
732	1	RHOA_HUMAN	193	2.8	5	805	1	CI30_ARATH	208	2.8	5	O96749 homo sapien
733	1	RHOA_MOUSE	193	2.8	5	806	1	FGF6_HUMAN	208	2.8	5	O9qui0 mus musculu
734	1	RPSH_PSEAE	193	2.8	5	807	1	FGF6_MOUSE	208	2.8	5	O06198 pseudomonas
735	1	FLHC_SERMA	194	2.8	5	808	1	IL6_BOVIN	208	2.8	5	O85807 serrattia ma
736	1	FLHC_XENNE	194	2.8	5	809	1	IL6_CAPHI	208	2.8	5	O9x9f3 xenorhabdus
737	1	RUVA_NEIMA	194	2.8	5	810	1	IL6_FELCA	208	2.8	5	O9j5m5 neisseria m
738	1	RUVA_NEIMA	194	2.8	5	811	1	IL6_HORSE	208	2.8	5	O9kia2 neisseria m
739	1	Y248_METJA	194	2.8	5	812	1	IL6_SHEEP	208	2.8	5	O57698 methanococc
740	1	CSLA_HUMAN	195	2.8	5	813	1	LEUD_ALCEU	208	2.8	5	O9y3b2 homo sapien
741	1	IGFB_HUMAN	195	2.8	5	814	1	NGN1_BRARE	208	2.8	5	P05019 homo sapien
742	1	INO1_HUMAN	195	2.8	5	815	1	NQRD_PASMU	208	2.8	5	P05000 homo sapien
743	1	ROAB_ARTSA	195	2.8	5	816	1	NU4M_MTCPE	208	2.8	5	P80350 artemia sal
744	1	YAGZ_ECOLI	195	2.8	5	817	1	RB12_CANFA	208	2.8	5	P77264 escherichia
745	1	KCY_HUMAN	196	2.8	5	818	1	UPP_AQUAE	208	2.8	5	P30085 homo sapien
746	1	KCY_PIG	196	2.8	5	819	1	UPP_ECOLI	208	2.8	5	O29561 sus scrofa
747	1	RNH2_AQUAE	196	2.8	5	820	1	UPP_HAEIN	208	2.8	5	O67768 aquifex aeo
748	1	YMO2_PORTE	196	2.8	5	821	1	UPP_NEIMA	208	2.8	5	P15604 paramedium
749	1	MOBA_STRECO	197	2.8	5	822	1	UPP_NEIMB	208	2.8	5	O86535 streptomyc
750	1	MOBA_METTH	197	2.8	5	823	1	UPP_PASMU	208	2.8	5	O26246 methanobact
751	1	R16B_YEAST	197	2.8	5	824	1	UPP_VIBCH	208	2.8	5	P26785 saccharomyc
752	1	RB12_RAT	197	2.8	5	825	1	YHNM_ECOLI	208	2.8	5	P35284 rattus norv
753	1	RL22_MYCTU	197	2.8	5	826	1	GTP1_MOUSE	209	2.8	5	P95054 mycobacteri
754	1	RUVA_BORBU	197	2.8	5	827	1	GTP2_MOUSE	209	2.8	5	P70827 borrelia bu
755	1	Y146_NPPOP	197	2.8	5	828	1	GTP_CRIMI	209	2.8	5	O10375 orgyia pseu
756	1	Y149_TREPA	197	2.8	5	829	1	IL6_PHOVI	209	2.8	5	O83184 treponema p
757	1	Y882_METJA	197	2.8	5	830	1	RL3_BACHD	209	2.8	5	O58292 methanococc
758	1	AMOS_DROME	198	2.8	5	831	1	UPP_BACSU	209	2.8	5	O9y0a7 drosophila
759	1	LEUD_MYCTU	198	2.8	5	832	1	UPP_BAGSU	209	2.8	5	O53236 mycobacteri
760	1	MOBA_PSEAE	198	2.8	5	833	1	UPP_LACPL	209	2.8	5	O68799 pseudomonas
761	1	R16A_YEAST	198	2.8	5	834	1	UPP_MYCHO	209	2.8	5	P26784 saccharomyc
762	1	PEXH_YEAST	199	2.8	5	835	1	UPP_THEMA	209	2.8	5	P40155 saccharomyc
763	1	TORD_ECO57	199	2.8	5	836	1	YBDM_ECOLI	209	2.8	5	P58364 escherichia

P36662 escherichia  
O58016 methanococc  
P58897 corynebacte  
P22264 xanthomonas  
P33883 pseudomonas  
O67034 aquifex aeo  
O8u345 agrobacteri  
O8yeg8 bruceella me  
O74895 schizosacch  
O97799 mycobacteri  
O06325 mycobacteri  
P23744 methylomona  
P30408 homo sapien  
O04905 arabidopsis  
Q15040 homo sapien  
P20538 vaccinia vl  
P07242 vaccinia vl  
O29208 archaeoglob  
P77929 pseudomonas  
P53654 xanthomonas  
P44272 haemophilus  
P26775 mus musculu  
P94614 coxiella bu  
O83723 treponema p  
Q55854 synechocyst  
Q28747 orcinus orc  
P01374 homo sapien  
O94655 schizosacch  
O08542 mus musculu  
P22492 homo sapien  
Q00630 manduca sex  
P58134 astasia lon  
Q9r773 deinococcus  
P40286 macaca mula  
P41323 canis famil  
Q35736 marmota mon  
Q12328 saccharomyc  
O50110 pyrococcus  
P28601 bacillus st  
P27672 coturnix co  
P94928 mycobacteri  
Q91q17 arabidopsis  
P10767 homo sapien  
P21658 mus musculu  
P26892 bos taurus  
Q28319 capra hircu  
P41683 felis salve  
Q95181 equus cabal  
P29455 ovis aries  
Q44022 alcaligenes  
O42606 brachydanio  
Q9cia8 pasteurella  
O21521 microtus pe  
P51152 canis famil  
O67914 aquifex aeo  
P25532 escherichia  
P43857 haemophilus  
Q9jv58 neisseria m  
Q9k048 neisseria m  
Q9cpl8 pasteurella  
Q9kpy7 vibrio chol  
P37616 escherichia  
P46425 mus musculu  
P19157 mus musculu  
P47954 cricetus  
Q28819 phoca vitul  
P42920 bacillus su  
Q9k6g5 bacillus ha  
P39149 bacillus su  
Q9re01 lactobacill  
P43049 mycoplasma  
Q9w210 thermotoga  
P77174 escherichia

837	1	HPRT_TRYBB	2.8	210	Q07010	trypanosoma	910	1	OAZ_RAT	226	1	P54370	rattus norv
838	1	PDX5_MOUSE	2.8	210	P99029	mus musculus	911	1	VATE_HFTSC	226	1	Q9ulq5	heterodera
839	1	UPP_DEIRA	2.8	210	Q9ru32	deinococcus	912	1	CBVY_RHOCA	227	1	Q33513	rhodobacter
840	1	NADD_MYCTU	2.8	211	O86328	mycobacteri	913	1	COX2_CHEME	227	1	P98027	cheirogaleu
841	1	PAAD_VIECH	2.8	211	Q9kp38	vibrio chol	914	1	COX2_EULMA	227	1	P98033	eulemur mac
842	1	IL6_CERTO	2.8	212	P46650	cercocoebus	915	1	COX2_GALSA	227	1	P50688	galago sene
843	1	IL6_HUMAN	2.8	212	P05231	homo sapien	916	1	COX2_HAPGR	227	1	P98034	hapalemur g
844	1	IL6_MACFA	2.8	212	P79341	macaca fasc	917	1	COX2_LBMVA	227	1	P98035	lemur catta
845	1	IL6_MACMU	2.8	212	P51494	macaca mula	918	1	COX2_LBMVA	227	1	P98047	lemur varie
846	1	IL6_PIG	2.8	212	P26893	sus scrofa	919	1	COX2_PROTH	227	1	P98042	proplithecus
847	1	RAN_DICDI	2.8	212	P33519	dictyosteli	920	1	RISB_ARATH	227	1	O80575	arabidopsis
848	1	ACD3_PSEAE	2.8	213	Q9hz17	pseudomonas	921	1	YMO_YEAST	227	1	O80575	arabidopsis
849	1	INL6_HUMAN	2.8	213	Q9y581	homo sapien	922	1	CLPR_SYN7	228	1	O04304	saccharomyc
850	1	PDX5_RAT	2.8	213	Q9r063	rattus norv	923	1	HS74_LEIMA	228	1	Q9144p	synchococc
851	1	PIT_STRHA	2.8	213	P41132	streptomyce	924	1	IF6_METJA	228	1	P12077	leishmania
852	1	RABB_DICDI	2.8	213	P34142	dictyosteli	925	1	MODE_RHOCA	228	1	O60357	methanococc
853	1	SKGR_XENLA	2.8	213	P13673	xenopus lae	926	1	Y516_BORBU	228	1	Q08382	rhodobacter
854	1	IYH2_YEAST	2.8	213	P40359	saccharomyc	927	1	Y972_METJA	228	1	O51468	borrellia bu
855	1	CMPA_BACSU	2.8	214	P14204	bacillus su	928	1	YR16_MYCTU	228	1	O58382	methanococc
856	1	GSI_HUMAN	2.8	214	Q08623	homo sapien	929	1	MTN_HABIN	229	1	O07215	mycobacteri
857	1	HS7A_DROSI	2.8	214	P02826	drosophila	930	1	PRL_MEIGA	229	1	P45113	haemophilus
858	1	MAPA_CAMJE	2.8	214	O46122	campylobact	931	1	RAS2_NEUCR	229	1	P17572	melagris g
859	1	NADD_MYCLE	2.8	214	Q9cbz8	mycobacteri	932	1	SFSA_THETN	229	1	Q01387	neurospora
860	1	PDX5_HUMAN	2.8	214	P30044	h peroxired	933	1	UCK_CAEEL	229	1	Q8r7n6	thermoanaer
861	1	PPCT_MOUSE	2.8	214	P53808	mus musculus	934	1	Y454_TREPA	229	1	Q17413	caenorhabdi
862	1	PPCT_RAT	2.8	214	P53809	rattus norv	935	1	YD74_MYCPN	229	1	O83467	treponema p
863	1	ERD2_ARATH	2.8	215	P35402	arabidopsis	936	1	AG11_ARATH	230	1	P75407	mycoplasma
864	1	HEX8_ADEML	2.8	215	P19722	mouse adeno	937	1	Y336_YEAST	230	1	Q38836	arabidopsis
865	1	KAD_MYCPN	2.8	215	Q50299	mycoplasma	938	1	YLM6_BACSU	230	1	P53274	saccharomyc
866	1	PDX5_CERAE	2.8	215	Q9glw7	cercopithe	939	1	YSHA_SALTY	230	1	O31727	bacillus su
867	1	PDX5_PAPHA	2.8	215	Q9glw9	papio hamad	940	1	NU4M_HYPHY	231	1	O917r3	salmonella
868	1	RANG_SCHPO	2.8	215	Q09717	schizosacch	941	1	NUC_STAAU	231	1	O03733	hypnale hyp
869	1	YHID_ECOLI	2.8	215	P26606	escherichia	942	1	PSA2_TRYBB	231	1	P00644	staphylococ
870	1	PUR7_AQUAE	2.8	216	O67881	aquifex aeo	943	1	Y493_RICPR	231	1	Q9u793	trypanosoma
871	1	UPP_PORGI	2.8	216	Q9zu68	porphyromon	944	1	F4RE_METTH	232	1	O92454	ricketsia
872	1	GRAN_HUMAN	2.8	217	P28676	homo sapien	945	1	GU58_RAT	232	1	O26350	methanobact
873	1	RL6_CAEEL	2.8	217	P47991	caenorhabdi	946	1	PHU1_THEMEA	232	1	P34899	rattus norv
874	1	Y038_METJA	2.8	217	O60345	methanococc	947	1	UBIG_PSEAE	232	1	Q9x0y7	thermotoga
875	1	B10D_HELPJ	2.8	218	Q9zu34	helicobacte	948	1	MINC_XYLFA	233	1	O9hz63	pseudomonas
876	1	B10D_HELPY	2.8	218	O24872	helicobacte	949	1	MIP_LEGLO	233	1	Q9pdq7	xylella fas
877	1	MSRA_BRUME	2.8	218	Q8yde7	brucella me	950	1	MIP_LEGPN	233	1	P53605	legionella
878	1	MSRA_OCHAN	2.8	218	Q93s39	ochrobactru	951	1	VU79_HSV7J	233	1	P20380	legionella
879	1	RPE_TREPA	2.8	218	O66107	treponema p	952	1	NOLW_RHIFR	234	1	P52531	human herpe
880	1	YBOF_SCHPO	2.8	218	P78860	schizosacch	953	1	NOLW_RHISN	234	1	P33212	rhizobium f
881	1	PDX5_BOVIN	2.8	219	Q9bdl1	bos taurus	954	1	RL6_MESCR	234	1	P55712	rhizobium s
882	1	RB3A_BOVIN	2.8	219	P11023	bos taurus	955	1	S120_YEAST	234	1	P34091	mesembryant
883	1	EXBB_NEIGO	2.8	220	O06433	neisseria g	956	1	SODM_CANAL	234	1	P39931	saccharomyc
884	1	EXBB_NEIMA	2.8	220	P57027	neisseria m	957	1	EFTU_DERMA	235	1	O13401	candida alb
885	1	EXBB_NEIMC	2.8	220	P95375	neisseria m	958	1	FL3L_HUMAN	235	1	P50375	derbesia ma
886	1	RB3A_MOUSE	2.8	220	P05713	mus musculus	959	1	PSAF_HORVU	235	1	P49771	homo sapien
887	1	GT21_DIACA	2.8	221	P28342	dianthus ca	960	1	TNFA_MOUSE	235	1	P13192	hordeum vul
888	1	BPBH_CAEEL	2.8	221	O16264	caenorhabdi	961	1	TNFA_RAT	235	1	P06804	mus musculus
889	1	RAN1_ARATH	2.8	221	P41916	arabidopsis	962	1	PUR7_PSEAE	236	1	P16599	rattus norv
890	1	YR75_CAEEL	2.8	221	Q09394	caenorhabdi	963	1	T650_BRUOV	236	1	Q914w0	pseudomonas
891	1	FLGH_RHOSH	2.8	222	Q28267	canis famil	964	1	Y385_MYCGE	236	1	Q08082	bruceella ov
892	1	I12A_CANFA	2.8	222	O02743	felis silve	965	1	Y385_MYCPN	237	1	P37425	mycoplasma
893	1	I12A_FELCA	2.8	222	Q9xsq6	equus cabal	966	1	PS52_ARATH	237	1	P47492	bacillus su
894	1	I12A_HORSE	2.8	222	Q42599	arabidopsis	967	1	PSA5_ORYSA	237	1	O81149	arabidopsis
895	1	NUIM_ARATH	2.8	222	Q927h0	chlamydia p	968	1	PSA5_ORYSA	237	1	O42134	arabidopsis
896	1	YC13_KLEPN	2.8	222	Q48459	klebsiella	969	1	PSA5_ORYSA	237	1	Q91su1	oryza sativ
897	1	HEX8_ADEP3	2.8	223	O83453	porcine ade	970	1	PUR7_BACHD	237	1	Q9mt48	glycine max
898	1	SODM_ONCVO	2.8	223	P41981	onchocerca	971	1	RPIA_AERPE	237	1	Q9kt60	bacillus ha
899	1	RNT_PSEAE	2.8	224	P26236	rhodobacter	972	1	Y385_MYCPN	237	1	Q9yea9	acropyrum p
900	1	BCNM_RHOCA	2.8	224	Q9hv82	pseudomonas	973	1	6PGL_PSEAE	238	1	P75212	mycoplasma
901	1	UOG_CHAFE	2.8	224	O96347	charybdis f	974	1	ATP6_MYTED	238	1	Q9x2n2	pseudomonas
902	1	UNG_BACHD	2.8	224	Q9k682	bacillus ha	975	1	GLO2_HAEIN	238	1	Q00224	mytilus edu
903	1	Y691_CHLTR	2.8	224	O84697	chlamydia t	976	1	PCXB_PSEPU	238	1	P71374	haemophilus
904	1	DEOC_SYNY3	2.8	225	P73618	synchocyst	977	1	Y834_METJA	238	1	P00437	pseudomonas
905	1	TN12_MOUSE	2.8	225	O54907	mus musculus	978	1	Y939_HAEIN	238	1	Q58244	methanococc
906	1	Y810_METJA	2.8	226	Q58220	methanococc	979	1	RLPA_NEIMA	239	1	P44080	haemophilus
907	1	MOBB_AZOVI	2.8	226	P37731	azotobacter	980	1	Y4BF_SCHPO	239	1	P58505	homo sapien
908	1	OAZ_MOUSE	2.8	226	P54369	mus musculus	981	1	Y4BF_SCHPO	239	1	Q9Jsm7	neisseria m
909	1	OAZ_MOUSE	2.8	226	P54369	mus musculus	982	1	Y4BF_SCHPO	239	1	P55491	rhizobium s
												Q09814	schizosacch



```
OC Escherichia.
OX NCBI_TaxID=562;
RP STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RN Science 277:1453-1474(1997).
RP STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RN DNA Res. 3:137-155(1996).
RP STRAIN=NM8191;
RX Ivey D.M., Guffanti A.A., Zemsky J., Pinner E., Karpel R.,
RA Padan E., Schuldiner S., Krulwich T.A.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN STRAIN=K12;
RX MEDLINE=89263708; PubMed=2657652;
RA Nohno T., Noji S., Taniguchi S., Saito T.;
RT "The narX and narL genes encoding the nitrate-sensing regulators of
RT Escherichia coli are homologous to a family of prokaryotic two-
RT component regulatory genes.";
RN Nucleic Acids Res. 17:2947-2957(1989).
RP STRAIN=K12;
RX MEDLINE=89197802; PubMed=2649492;
RA Stewart V., Parates J. Jr., Merkel S.M.;
RT "Structure of genes narL and narX of the nar (nitrate reductase)
RT locus in Escherichia coli K-12.";
RL J. Bacteriol. 171:2229-2234(1989).
RP IDENTIFICATION.
RX MEDLINE=96032851; PubMed=7567469;
RA Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,
RA Danchin A.;
RT "Detection of new genes in a bacterial genome using Markov models for
RT three gene classes.";
RL Nucleic Acids Res. 23:3554-3562(1995).
CC -1- SIMILARITY: BELONGS TO THE INTIMIN/INVASIN FAMILY.
CC -1- CAUTION: REF.4 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 398.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000220; AAC74304.1; -
CC DR EMBL; D90757; BAA36088.1; -
CC DR EMBL; L28709; -; NOT_ANNOTATED_CDS.
CC DR EMBL; X13360; -; NOT_ANNOTATED_CDS.
CC DR EMBL; M24910; -; NOT_ANNOTATED_CDS.
CC -----
CC Query Match 3.9%; Score 7; DB 1; Length 146;
CC Best Local Similarity 100.0%; Pred. No. 7.7;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;
CC -----
QY 158 VEADYTT 164
DB 13 VEADYTT 19
RESULT 5
MYG_THUAL
```

```
DR EMBL; X69189; -; NOT_ANNOTATED_CDS.
DR EcoGene; EG12405; ycho.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 417 AA; 47023 MW; 4D4841DDC1BE5485 CRC64;
Query Match 4.5%; Score 8; DB 1; Length 417;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Caps 0;
QY 172 GLNLNRYRF 179
DB 256 GLNLNRYRF 263
RESULT 4
MYG_GORGI
ID MYG_GORGI STANDARD; PRT; 146 AA.
AC Q9DEN9;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myoglobin.
GN Mb.
OS Gobionotothen gibberifrons (Humped rockcod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes;
OC Notothenioidae; Nototheniidae; Gobionotothen.
OX NCBI_TaxID=36202;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart ventricle;
RX MEDLINE=97430285; PubMed=9284539;
RA Vayda M.E., Small D.J., Yuan M.-L., Costello L., Sidell B.D.;
RT "Conservation of the myoglobin gene among Antarctic notothenioid
RT fishes.";
RL Mol. Mar. Biol. Biotechnol. 6:207-216(1997).
CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES
CC THE MOVEMENT OF OXYGEN WITHIN MUSCLES.
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U71057; AAG16645.1; -
CC DR HSSP; P02205; 1MYT.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 2
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT INIT_MET 0 0 BY SIMILARITY.
FT METAL 59 59 IRON (HEME DISTAL LIGAND).
FT METAL 88 88 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 146 AA; 15576 MW; 6DAAEF3DEB3CE109 CRC64;
Query Match 3.9%; Score 7; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;
QY 158 VEADYTT 164
DB 13 VEADYTT 19
RESULT 5
MYG_THUAL
```

```

ID MYG_THUAL STANDARD; PRT; 146 AA.
AC P02205;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myoglobin.
GN MB.
OS Thunnus albacares (Yellowfin tuna) (Neothunnus macropterus).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Scombroidei;
OC Scombridae; Thunnus.
OX NCBI_TaxID=8236;
RN [1]
RP SEQUENCE.
RX MEDLINE=81046956; PubMed=7430163;
RA Watts D.A., Rice R.H., Brown W.D.;
RT "The primary structure of myoglobin from yellowfin tuna (Thunnus
  albacares).";
RL J. Biol. Chem. 255:10916-10924(1980).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.74 ANGSTROMS).
RA Birnbaum G.I., Evans S.V., Przybylska M., Rose D.R.;
RT "1.70-A resolution structure of myoglobin from yellowfin tuna. An
  example of a myoglobin lacking the D helix.";
RL Acta Crystallogr. D 50:283-289(1994).
CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES
  THE MOVEMENT OF OXYGEN WITHIN MUSCLES.
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
DR PIR; A02526; MTUY.
DR PDB; 1MYT; 3I-OCT-93.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle; Acetylation;
  3D-structure.
FT MOD_RES 1 1
FT METAL 59 59 ACETYLATION.
FT FT METAL 59 59 IRON (HEME DISTAL LIGAND)
FT FT METAL 88 88 (BY SIMILARITY).
FT FT METAL 88 88 IRON (HEME PROXIMAL LIGAND)
FT FT METAL 88 88 (BY SIMILARITY).
FT FT HELIX 2 8
FT FT TURN 9 10
FT FT HELIX 11 14
FT FT TURN 15 16
FT FT HELIX 17 31
FT FT HELIX 33 38
FT FT TURN 40 44
FT FT HELIX 47 50
FT FT TURN 51 52
FT FT HELIX 54 72
FT FT TURN 73 75
FT FT HELIX 78 90
FT FT TURN 91 91
FT FT TURN 96 96
FT FT HELIX 97 113
FT FT HELIX 118 142
FT FT TURN 143 143
SQ SEQUENCE 146 AA; 15529 MW; 74C72B21D005BD44 CRC64;

Query Match 3.9%; Score 7; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 VEADYTT 164
DB 13 VEADYTT 19

RESULT 6
MYG_THUTH

```

```

ID MYG_THUTH STANDARD; PRT; 146 AA.
AC Q9DD47;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myoglobin.
GN MB.
OS Thunnus thynnus (Bluefin tuna), and
  Thunnus thynnus orientalis (North Pacific bluefin tuna).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Scombroidei;
OC Scombridae; Thunnus.
OX NCBI_TaxID=8237, 8238;
RN [1]
RP SEQUENCE FROM N.A.
RX SPECIES=T.thynnus, and T.t.orientalis; TISSUE=Skeletal muscle;
  MEDLINE=21146139; PubMed=11247835;
RA Marcinek D.J., Bonaventura J., Wittenberg J.B., Block B.A.;
RT "Oxygen affinity and amino acid sequence of myoglobins from
  endothermic and ectothermic fish.";
RL Am. J. Physiol. 280:R1123-R1133(2001).
CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES
  THE MOVEMENT OF OXYGEN WITHIN MUSCLES.
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF291836; AAC02110.1;
DR EMBL; AF291831; AAC02105.1;
DR HSSP; P02205; 1MYT.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 2.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT INIT_MET 0 0 BY SIMILARITY.
FT METAL 59 59 IRON (HEME DISTAL LIGAND).
FT METAL 88 88 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 146 AA; 15497 MW; F99735CEC834509A CRC64;

Query Match 3.9%; Score 7; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 VEADYTT 164
DB 13 VEADYTT 19

RESULT 7
CAH2_MOUSE
ID CAH2_MOUSE STANDARD; PRT; 259 AA.
AC P00920;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carbonic anhydrase II (EC 4.2.1.1) (Carbonate dehydratase II) (CA-II).
GN CA2 OR CAR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84109569; PubMed=6420240;

```

RA Curtiss P.J., Withers E., Demuth D., Watt R., Venta P.J., Tashian R.E.;  
 RT "The nucleotide sequence and derived amino acid sequence of cDNA  
 RL coding for mouse carbonic anhydrase II.";  
 RN Gene 25:325-332(1983).

RP SEQUENCE FROM N.A.  
 RX MEDLINE=86008276; PubMed=2995362;  
 RA Venta P.J., Montgomery J.C., Hewett-Emmett D., Wiebauer K.,  
 RA Tashian R.E.;  
 RT "Structure and exon to protein domain relationships of the mouse  
 RT carbonic anhydrase II gene.";  
 RN J. Biol. Chem. 260:12130-12135(1985).

RP SEQUENCE OF 154-177 AND 213-239 FROM N.A.  
 RX MEDLINE=83161023; PubMed=6187736;  
 RA Curtiss P.J.;  
 RT "Cloning of mouse carbonic anhydrase mRNA and its induction in mouse  
 RT erythroleukemic cells.";  
 RN J. Biol. Chem. 258:4459-4463(1983).

RP SEQUENCE OF 1-76 FROM N.A.  
 RX MEDLINE=86077780; PubMed=3000449;  
 RA Venta P.J., Montgomery C., Hewett-Emmett D., Tashian R.E.;  
 RT "Comparison of the 5' regions of human and mouse carbonic anhydrase  
 RT II genes and identification of possible regulatory elements.";  
 RL Biochim. Biophys. Acta 826:195-201(1985).

CC -!- FUNCTION: REVERSIBLE HYDRATATION OF CARBON DIOXIDE.  
 CC -!- CATALYTIC ACTIVITY: H<sub>2</sub>CO<sub>3</sub> = CO<sub>2</sub> + H<sub>2</sub>O.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: BELONGS TO THE EUKARYOTIC-TYPE CARBONIC ANHYDRASE  
 CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; K00811; AAA37356.1;  
 DR EMBL; M81022; AAA37357.1;  
 DR EMBL; M81016; AAA37357.1; JOINED.  
 DR EMBL; M81017; AAA37357.1; JOINED.  
 DR EMBL; M81018; AAA37357.1; JOINED.  
 DR EMBL; M81019; AAA37357.1; JOINED.  
 DR EMBL; M81020; AAA37357.1; JOINED.  
 DR EMBL; M81021; AAA37357.1; JOINED.  
 DR PIR; A01143; CRMS2.  
 DR PIR; A23900; A23900.  
 DR PIR; A20539; A20539.  
 DR PIR; B23202; B23202.  
 DR HSSP; P00918; 1BV3.  
 DR MGD; MGI:88269; Car2.  
 DR InterPro; IPR001148; Euk\_Coanhd.  
 DR Pfam; PF00194; carb\_anhydrase; 1.  
 DR ProDom; PD000865; Euk\_Coanhd; 1.  
 DR PROSITE; PS00162; Euk\_CO2\_ANHYDRASE; 1.

FW Lyase; Zinc.  
 FT INIT\_MET 0 0  
 FT METAL 93 93 ZINC (CATALYTIC).  
 FT METAL 95 95 ZINC (CATALYTIC).  
 FT METAL 118 118 ZINC (CATALYTIC).  
 FT METAL 38 38 H -> Q (IN REF. 2).  
 FT CONFLICT 212 212 E -> D (IN REF. 2).  
 SQ SEQUENCE 259 AA; 28960 MW; 5BD81AD512D3A86B CRC64;

Query Match 3.9%; Score 7; DB 1; Length 259;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 YDKAASK 103  
 |||||

DB 50 YDKAASK 56  
 RESULT 8  
 ID GPDA\_STRCO STANDARD; PRT; 336 AA.  
 AC Q92BS0;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Glycerol-3-phosphate dehydrogenase [NAD(P)+] (EC 1.1.1.94) (NAD(P)H-  
 DE dependent glycerol-3-phosphate dehydrogenase).  
 GN GPSA OR SC05559 OR SC7A1.03.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;  
 OC Actinomycetales; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 CC -!- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(P)+ =  
 CC glycerone phosphate + NAD(P)H.  
 CC -!- PATHWAY: De novo phospholipid biosynthesis; glycerol-3 phosphate  
 CC formation.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -!- SIMILARITY: BELONGS TO THE NAD-DEPENDENT GLYCEROL-3-PHOSPHATE  
 CC DEHYDROGENASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AL034447; CAA22402.1;  
 DR InterPro; IPR001652; NAD\_Gly3p\_dh.  
 DR Pfam; PF01210; NAD\_Gly3p\_dh; 1.  
 DR PRINTS; PR00077; GPDHGRGNASE.  
 DR ProDom; PD001649; NAD\_Gly3p\_dh; 1.  
 DR PROSITE; PS00957; NAD\_G3PDH; 1.  
 KW Phospholipid biosynthesis; Oxidoreductase; NAD; Complete proteome.  
 SQ SEQUENCE 336 AA; 34688 MW; 4E7BF65F9C640544 CRC64;

Query Match 3.9%; Score 7; DB 1; Length 336;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 TLRAGLA 96  
 |||||  
 DB 86 TLRAGLA 92

RESULT 9  
 ID CMGA\_BACSU STANDARD; PRT; 356 AA.  
 AC P25953;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)

15-JUN-2002 (Rel. 41, Last annotation update)  
 COMG operon protein 1.  
 GN COMGA OR COMG1.  
 OS Bacillus subtilis.  
 OC Bacteria: Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90008773; PubMed=2507524;  
 RA Albano M., Breitling R., Dubnau D.A.;  
 RT "Nucleotide sequence and genetic organization of the Bacillus  
 subtilis comG operon."  
 RL J. Bacteriol. 171:5386-5404(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / JH642;  
 RX MEDLINE=97124195; PubMed=8969508;  
 RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,  
 RA Kobayashi Y.;  
 RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of  
 the Bacillus subtilis genome containing the skin element and many  
 sporulation genes."  
 RL Microbiology 142:3103-3111(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borillas R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
 RA Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Giuseppe G., Guy B.J., Haga K., Hatach J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzengger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 subtilis."  
 RL Nature 390:249-256(1997).  
 RN [4]  
 RP FUNCTION.  
 RX MEDLINE=98083053; PubMed=9422590;  
 RA Chung Y.S., Dubnau D.A.;  
 RT "All seven comG open reading frames are required for DNA binding  
 during transformation of competent Bacillus subtilis."  
 RL J. Bacteriol. 180:41-45(1998).  
 CC -1- FUNCTION: REQUIRED FOR UPTAKE OF DNA BY COMPETENT BACILLUS  
 CC SUBTILIS CELLS.  
 CC -1- SUBCELLULAR LOCATION: INNER FACE OF THE MEMBRANE.  
 CC -1- SIMILARITY: BELONGS TO THE PULE/OUTE/EXEE/XPCR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 -----  
 DR EMBL; M29691; AAA83367.1; -;  
 DR EMBL; D84432; BAA12533.1; -;  
 DR EMBL; Z99116; CAB14404.1; -;  
 DR PTR; B30338; B30338;  
 DR Subtilist; BG10483; comGA.  
 DR InterPro; IPR001482; GSPIL\_E.  
 DR Pfam; PF00437; GSPIL\_E; 1.  
 DR ProDom; PD000739; GSPIL\_E; 1.  
 DR PROSITE; PS00662; T2SP\_E; 1.  
 DR KW Transport; ATP-binding; Complete proteome.  
 FT NP\_BIND 144 151 ATP (POTENTIAL).  
 SQ SEQUENCE 356 AA; 40458 MW; C15840FFB734E266 CRC64;  
 Query Match 3.9%; Score 7; DB 1; Length 356;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 13 GSLTLKL 19  
 Db 77 GSLTLKL 83  
 -----  
 RESULT 10  
 HN3B\_ORYLA STANDARD; PRT; 415 AA.  
 ID HN3B\_ORYLA  
 AC 042097;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hepatocyte nuclear factor 3-beta (HNF-3b) (Me-HNF3B).  
 GN HNF3B.  
 OS Oryzias latipes (Medaka fish) (Japanese ricefish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
 OC Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.  
 OX NCBI\_TaxID=8090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Orange-red;  
 RA Okamoto H.M., Nakayama I., Nagoya H., Araki K.;  
 RT "Predicted protein structure of medaka FoxA3 and its expression in  
 Zool. Sci. 18:823-832(2001).  
 RL polster";  
 CC -1- FUNCTION: TRANSCRIPTION ACTIVATOR FOR A NUMBER OF LIVER GENES.  
 CC INTERACTS WITH THE CIS-ACTING REGULATORY REGIONS OF THESE GENES.  
 CC (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 -----  
 DR EMBL; AB001572; BAA23579.1; -;  
 DR HSSP; Q63245; 2HFH.  
 DR InterPro; IPR001766; TF\_Fork\_head.  
 DR Pfam; PF00250; Fork\_head; 1.  
 DR PRINTS; PR00053; FORKHEAD.  
 DR ProDom; PD000425; TF\_Fork\_head; 1.  
 DR SMART; SM00339; FH; 1.  
 DR PROSITE; PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE; PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE; PS00039; FORK\_HEAD\_3; 1.



KW DNA-binding: Nuclear protein; Transcription regulation; Activator.  
 FT DOMAIN 32 124 MET-RICH.  
 FT DNA\_BIND 150 244 FORK-HEAD.  
 SQ SEQUENCE 415 AA; 45722 MW; CB85A0C2A20FF52A CRC64;

Query Match 3.9%; Score 7; DB 1; Length 415;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 SASIPDT 113  
 |||||

Db 394 SASIPDT 400

# RESULT 11

OM47\_PASMU STANDARD; PRT; 428 AA.  
 AC P80603; Q9CLOY; 428 AA.  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 47 kDa outer membrane protein precursor.  
 GN Pml069.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Pasteurella.  
 OX NCBI\_TaxID=747;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Pm70;  
 RX MEDLINE=21145866; PubMed=11248100;  
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
 RT "Complete genomic sequence of Pasteurella multocida Pm70.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
 RN [2]  
 RP SEQUENCE OF 26-44.  
 RC STRAIN=A225;  
 RA Hartmann L.;  
 RL Submitted (MAY-1996) to the SWISS-PROT data bank.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.  
 CC -1- SIMILARITY: BELONGS TO THE OMP1/FADL FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; A5006147; AAK03153.1; ALT\_INIT.  
 CC Outer membrane; Signal; Transmembrane; Complete proteome.  
 KW SIGNAL 1 25  
 FT CHAIN 26 428 47 KDa OUTER MEMBRANE PROTEIN.  
 SQ SEQUENCE 428 AA; 45061 MW; D9606AC50F8220EC CRC64;  
 Query Match 3.9%; Score 7; DB 1; Length 428;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 HYSYKVT 44  
 |||||

Db 298 HYSYKVT 304

# RESULT 12

FADL\_ECOLI STANDARD; PRT; 448 AA.  
 ID FADL\_ECOLI  
 AC P10384; P77697;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Long-chain fatty acid transport protein precursor (Outer membrane

DE FADL protein) (Outer membrane flp protein).  
 GN FADL OR TTR OR B2344.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-38.  
 RX MEDLINE=91100327; PubMed=1987139;  
 RA Black P.N.;  
 RT "Primary sequence of the Escherichia coli fadL gene encoding an outer  
 RT membrane protein required for long-chain fatty acid transport.";  
 RL J. Bacteriol. 173:435-442(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick J.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97349980; PubMed=9205837;  
 RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,  
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,  
 RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,  
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
 RA Yamagata S., Horiuchi T.;  
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
 RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and  
 RT analysis of its sequence features.";  
 RL DNA Res. 4:91-113(1997).  
 RN [4]  
 RP SEQUENCE OF 66-448 FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=88288050; PubMed=2840553;  
 RA Said B., Ghosh C.R., Vu L., Nunn W.;  
 RT "Nucleotide sequencing and expression of the fadL gene involved in  
 RT long-chain fatty acid transport in Escherichia coli.";  
 RL Mol. Microbiol. 2:363-370(1988).  
 CC -1- FUNCTION: INVOLVED IN TRANSLOCATION OF LONG-CHAIN FATTY ACIDS  
 CC ACROSS THE OUTER MEMBRANE. IT IS A RECEPTOR FOR THE BACTERIOPHAGE  
 CC T2. FADL MAY FORM A SPECIFIC CHANNEL.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.  
 CC CONTAINS SEVERAL POTENTIAL TRANSMEMBRANE DOMAINS.  
 CC -1- INDUCTION: BY LONG-CHAIN FATTY ACIDS. EXPRESSION OF FADL IS UNDER  
 CC THE CONTROL OF THE FADR REPRESSOR.  
 CC -1- SIMILARITY: BELONGS TO THE OMP1/FADL FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M60607; AAA64433.1; -;  
 CC EMBL; A5000323; AAC75404.1; -;  
 CC EMBL; D90865; BAA16205.1; -;  
 CC EMBL; Y00552; CAA68630.1; -;  
 CC PIR; A39126; A39126.  
 CC PIR; S02829; S02829.  
 CC EcoGene; EG10280; fadL.  
 KW Outer membrane; Lipid transport; Phage recognition; Signal;  
 KW Transmembrane; Complete proteome.  
 FT SIGNAL 1 27  
 FT CHAIN 28 448 LONG-CHAIN FATTY ACID TRANSPORT PROTEIN.

```

FT CONFLICT 178 178 A -> V (IN REF. 1).
FT CONFLICT 180 180 ARA -> RRP (IN REF. 4).
FT CONFLICT 288 288 R -> A (IN REF. 2 AND 3).
SQ SEQUENCE 448 AA; 48857 MW; BIA48EA07210FA88 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 20; Length 448;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 QWAIHYS 40
Db 313 QWAIHYS 319
|||||||

RESULT 13
AMID_PSEPU STANDARD; PRT; 466 AA.
AC 069768;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Amidase (EC 3.5.1.4).
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN SEQUENCE FROM N.A.
RP STRAIN=NRRL 18668;
RX MEDLINE=99025625; PubMed=9809753;
RA Wu S., Fallon R.D., Payne M.S.;
RT "Cloning and nucleotide sequence of amidase gene from Pseudomonas putida.";
RL DNA Cell Biol. 17:915-920(1998).
CC -1- CATALYTIC ACTIVITY: A monocarboxylic acid amide + H(2)O -> a
monocarboxylate + NH(3).
CC -1- PATHWAY: INVOLVED IN THE METABOLISM OF NITRILES.
CC -1- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; U89363; AAC18422.1; -.
CC InterPro: IPR000120; Amidase.
CC Pfam; PF01425; Amidase; 1.
CC PROSITE; PS00571; AMIDASES; 1.
KW Hydrolyase.
SQ SEQUENCE 466 AA; 50162 MW; AB9FF6F10B7CF4CF5 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 21; Length 466;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 LAGLAY 97
Db 443 LAGLAY 449
|||||||

RESULT 14
CD19_MOUSE STANDARD; PRT; 547 AA.
ID CD19_MOUSE
AC P25918;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE B-lymphocyte antigen CD19 precursor (B-lymphocyte surface antigen
DE B4) (Leu-12) (Differentiation antigen CD19).
GN CD19.

```

```

OS Mus musculus. (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN SEQUENCE FROM N.A.
RP STRAIN=BALB/c; TISSUE=Spleen;
RX MEDLINE=9132460; PubMed=1714482;
RA Zhou L.J., Ord D.C., Hughes A.L., Tedder T.F.;
RT "Structure and domain organization of the CD19 antigen of human,
RT mouse, and guinea pig B lymphocytes. Conservation of the extensive
RT cytoplasmic domain.";
RL J. Immunol. 147:1424-1432(1991).
RN SEQUENCE FROM N.A.
RP TISSUE=Spleen;
RX MEDLINE=92137946; PubMed=1370948;
RA Zhou L.J., Ord D.C., Omori S.A., Tedder T.F.;
RT "Structure of the genes encoding the CD19 antigen of human and mouse
RT B lymphocytes.";
RL Immunogenetics 35:102-111(1992).
RN SEQUENCE OF 302-547 FROM N.A.
RP STRAIN=BALB/c; TISSUE=Spleen;
RX MEDLINE=89292699; PubMed=2472450;
RA Tedder T.F., Isaacs C.M.;
RT "Isolation of cDNAs encoding the CD19 antigen of human and mouse B
RT lymphocytes. A new member of the immunoglobulin superfamily.";
RL J. Immunol. 143:712-717(1989).
RN SEQUENCE OF 1-26 FROM N.A.
RP MEDLINE=92269839; PubMed=1375324;
RA Kozmik Z., Wang S., Doerfler P., Adams B., Busslinger M.;
RT "The promoter of the CD19 gene is a target for the B-cell-specific
RT transcription factor BSAP.";
RL Mol. Cell. Biol. 12:2662-2672(1992).
CC -1- FUNCTION: MAY BE INVOLVED IN GROWTH REGULATION OF B-CELLS.
CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; M62542; AAA37388.1; -.
CC EMBL; M62553; AAA37390.1; -.
CC EMBL; M62551; AAA37390.1; JOINED.
CC EMBL; M62552; AAA37390.1; JOINED.
CC EMBL; M28240; AAA74753.1; -.
CC EMBL; M84372; AAA37389.1; -.
CC PIR; B45808; B45808.
CC MGD; MGI:88319; Cd19.
CC InterPro: IPR003599; Ig.
CC InterPro: IPR003006; Ig_MHC.
CC Pfam; PF00047; Ig; 2.
CC SMART; SM00409; Ig; 2.
KW Immunoglobulin domain; Glycoprotein; Transmembrane; B-cell; Signal;
Repeat.
FT SIGNAL 1 16
FT PROPEP 17 18
FT CHAIN 19 547
FT DOMAIN 19 287
FT TRANSMEM 288 311
FT DOMAIN 312 547
FT DOMAIN 31 104
FT DOMAIN 190 266
FT DISULFID 38 97
FT DISULFID 197 259

```

```

FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .)
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 263 263 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 207 207 A -> G (IN REF. 1).
SQ SEQUENCE 547 AA; 60163 MW; 06256FDA1BC2CBF3 CRC64;

Query Match 3.9%; Score 7; DB 1; Length 547;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 QDGSYGE 63
DB 427 QDGSYGE 433

RESULT 15
CD19_HUMAN
ID CD19_HUMAN STANDARD; PRT; 556 AA.
AC P15391.
DT 01-APR-1990 (Rel. 14, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE B-lymphocyte antigen CD19 precursor (B-lymphocyte surface antigen
DE B4) (Leu-12) (Differentiation antigen CD19).
GN CD19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Tonsil.
RX MEDLINE=89292699; PubMed=2472450;
RA Tedder T.F., Isaacs C.M.;
RT "Isolation of cDNAs encoding the CD19 antigen of human and mouse B
RT lymphocytes. A new member of the immunoglobulin superfamily.";
RL J. Immunol. 143:712-717(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89010509; PubMed=2459292;
RA Stamenkovic I., Seed B.;
RT "CD19, the earliest differentiation antigen of the B cell lineage,
RT bears three extracellular immunoglobulin-like domains and an
RT Epstein-Barr virus-related cytoplasmic tail.";
RL J. Exp. Med. 168:1205-1210(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92269839; PubMed=1375324;
RA Kozmik Z., Wang S., Doerfler P., Adams B., Busslinger M.;
RT "The promoter of the CD19 gene is a target for the B-cell-specific
RT transcription factor BSAP.";
RL Mol. Cell. Biol. 12:2662-2672(1992).
RN [4]
RP SEQUENCE FROM N.A.
RX TISSUE=Blood;
RX MEDLINE=92137946; PubMed=1370948;
RA Zhou L.J., Ord D.C., Omori S.A., Tedder T.F.;
RT "Structure of the genes encoding the CD19 antigen of human and mouse
RT B lymphocytes.";
RL Immunogenetics 35:102-111(1992).
CC -!- FUNCTION: MAY BE INVOLVED IN GROWTH REGULATION OF B-CELLS.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -!- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 396
CC ONWARD AND IS SHORTER (467 AA) DUE TO A FRAMESHIFT.
CC -!- DATABASE: NAME=PROV; NOTE=CD guide CD19 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd19.htm".
-----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M28170; AAC68490.1; -
DR EMBL: M21097; AAA35533.1; ALT_FRAME.
DR EMBL: M84371; AAA69966.1; -
DR EMBL: M62550; AAB60697.1; -
DR EMBL: M62544; AAB60697.1; JOINED.
DR EMBL: M62545; AAB60697.1; JOINED.
DR EMBL: M62546; AAB60697.1; JOINED.
DR EMBL: M62547; AAB60697.1; JOINED.
DR EMBL: M62548; AAB60697.1; JOINED.
DR EMBL: M62549; AAB60697.1; JOINED.
DR PIR: JL0074; JL0074.
DR PIR: A45808; A45808.
DR PIR: A4441; A4441.
DR Genew: HGNC:1633; CD19.
DR MIM: 107265; -
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_2.
DR SMART: SM00409; Ig_2.
KW Immunoglobulin domain; Glycoprotein; Transmembrane; B-cell; Signal;
KW Repeat.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 556 B-LYMPHOCYTE ANTIGEN CD19.
FT DOMAIN 20 291 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 292 313 POTENTIAL.
FT DOMAIN 314 556 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 31 104 IG-LIKE C2-TYPE DOMAIN 1.
FT DOMAIN 193 268 IG-LIKE C2-TYPE DOMAIN 2.
FT DISULFID 38 97 POTENTIAL.
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 265 265 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 29 29 E -> EG (IN REF. 4).
FT CONFLICT 80 80 S -> I (IN REF. 2).
FT CONFLICT 186 186 Q -> QAEVLVSLPVP (IN REF. 3).
FT CONFLICT 514 514 H -> R (IN REF. 3 AND 4).
SQ SEQUENCE 556 AA; 61069 MW; 5957952A169AFB42 CRC64;

Query Match 3.9%; Score 7; DB 1; Length 556;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 QDGSYGE 63
DB 434 QDGSYGE 440

RESULT 16
FIBP_ADE31
ID FIBP_ADE31 STANDARD; PRT; 556 AA.
AC P36848;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Fiber protein.
GN FIB.
OS Human adenovirus type 31.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=10529;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VRL 15/62;

```

```

RX MEDLINE-96151051; PubMed=8578008;
RA Pring-Akerblom P., Adrian T.;
RT "Sequence characterization of the adenovirus 31 fibre and comparison
RL with serotypes of subgenera A to F.";
RS. Virol. 146:343-354(1995).
CC -1- FUNCTION: RECOGNIZES THE CELL RECEPTOR; SERVES AS THE LIGAND
CC BETWEEN THE ADENOVIRUS CAPSID AND THE HOST CELL RECEPTOR.
CC -1- SUBUNIT: HOMOTRIMER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X76548; CAA54050.1; -
CC FIRM: S39297; S39297.
CC HSP: P11818; IKNB.
CC InterPro: IPR000939; Adeno_fiber2.
CC InterPro: IPR000978; Adeno_fiber_knob.
CC InterPro: IPR000931; Adeno_fibre.
CC Pfam: PF00541; adeno_fiber.1.
CC Pfam: PF00608; adeno_fiber2.6.
CC PRINTS: PR00307; ADENOVSFIBRE.
CC Fiber protein.
KW MESH: 556 AA; 58944 MW; 6FB13A60ED1D0C33 CRC64;
SQ SEQUENCE 556 AA; 58944 MW; 6FB13A60ED1D0C33 CRC64;

Query Match 3.9%; Score 7; DB 1; Length 556;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 GSLTLKL 19
DB 66 GSLTLKL 72

RESULT 17
RPA2_METJA STANDARD; PRT; 859 AA.
ID RPA2_METJA STANDARD; PRT; 859 AA.
AC Q58446;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase subunit A* (EC 2.7.7.6) [Contains: Mja
DE rpoA2 intein (Mja rpol A* intein)].
DE RPOA2 OR MJ1043
GN Methanococcus jannaschii.
OS Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE-96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC [RNA](N).
CC -1- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES

```

```

CC A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
CC FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC -1- SIMILARITY: IN THE INTEIN SECTION; BELONGS TO THE HOMING
CC ENDONUCLEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67547; AAB99047.1; -
CC TIGR: MJ1043;
CC InterPro: IPR003586; Hedgehog_hintc.
CC InterPro: IPR003587; Hedgehog_hintcN.
CC InterPro: IPR002203; Intein.
CC InterPro: IPR004042; Intein_endonuc.
CC InterPro: IPR002879; RNA_pol_A2.
CC Pfam: PF01854; RNA_pol_A2.2.
CC SMART: SM00305; Hintc.1.
CC SMART: SM00306; HintcN.1.
CC PROSITE: PS50818; INTEIN_CTER.1.
CC PROSITE: PS50819; INTEIN_ENDONUCLEASE.1.
CC PROSITE: PS50817; INTEIN_NTER.1.
CC Transferrase; Transcription; DNA-directed RNA polymerase; Zinc;
KW Autocatalytic cleavage; Protein splicing; Complete proteome;
KW Hydrolase; Nuclease; 75 RPOA2, 1ST PART (POTENTIAL).
FT CHAIN 1 76 546 MJA RPOA2 INTEIN (SPACER PROTEIN)
FT CHAIN 547 859 (POTENTIAL).
FT CHAIN RPOA2, 2ND PART (POTENTIAL).
SQ SEQUENCE 859 AA; 97086 MW; A05799E007899015 CRC64;

Query Match 3.9%; Score 7; DB 1; Length 859;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 KEVKYKDN 74
DB 588 KEVKYKDN 594

RESULT 18
HTRE_ECOLI STANDARD; PRT; 865 AA.
ID HTRE_ECOLI STANDARD; PRT; 865 AA.
AC P33129;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane usher protein htrE precursor (Heat shock protein E).
DE HTRE OR B0139.
GN Escherichia coli.
OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-93352405; PubMed=8102362;
RA Raina S., Missiakos D., Baird L., Kumar S., Georgopoulos C.;
RT "Identification and transcriptional analysis of the Escherichia coli
RT htrE operon which is homologous to pap and related pilin operons.";
RL J. Bacteriol. 175:5009-5021(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / W3110;
RX MEDLINE-94261430; PubMed=8202364;
RA Fujita N., Mori H., Yura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 2.4-4.1 min (110,917-193,643 bp) region.";

```

Nucleic Acids Res. 22:1637-1639(1994).

[3] SEQUENCE FROM N.A.

RL STRAIN-K12 / MG1655;

RN MEDLINE-97426617; PubMed-9278503;

RC Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

RN [4]

RN SEQUENCE OF 30-33.

RA Raina S.;

RL Submitted (DEC-1994) to the SWISS-PROT data bank.

CC -1- FUNCTION: PROBABLE PORIN-LIKE PROTEIN NECESSARY FOR THE ASSEMBLY

CC OF A PILIN-TYPE PROTEIN.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane

CC (by similarity).

CC -1- INDUCTION: BY HEAT SHOCK.

CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.

CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A

CC FRAMESHIFT IN POSITION 481 WHICH CREATES TWO SEPARATE ORFS.

CC -1- CAUTION: REF.1 SEQUENCE WAS INCORRECT IN POSITION 861 ONWARD DUE

CC TO CLONING ARTIFACT.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL: L00680; AAA23721.1; ALT\_SEQ.

DR EMBL: D26562; -; NOT\_ANNOTATED\_CDS.

DR EMBL: AE000123; AAC73250.1; -.

DR EcoGene: EGI1972; HtrE.

DR InterPro: IPR000015; Fimb\_usher.

DR Pfam: PF00577; Usher; 1.

DR PROSITE: PS01151; FIMBRIAL\_USHER; 1.

DR Outer membrane; Transmembrane; Fimbriae; Transport; Signal; Heat shock;

KW Complete proteome.

FT SIGNAL 1 29

FT CHAIN 30 865 OUTER MEMBRANE USHER PROTEIN HTR E.

FT DISULFID 838 862 POTENTIAL.

FT CONFLICT 51 51 S -> T (IN REF. 1).

FT CONFLICT 810 811 OG -> HR (IN REF. 1).

FT CONFLICT 849 849 E -> P (IN REF. 1).

FT SEQUENCE 865 AA; 95499 MW; F1748B551E45AAE CRC64;

-----

Query Match 3.9%; Score 7; DB 1; Length 865;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 LNDALTL 91

Db 474 LNDALTL 480

-----

RESULT 19

A2AL\_MOUSE STANDARD; PRT; 977 AA.

ID A2AL\_MOUSE

AC P17426;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Adapter-related protein complex 2 alpha 1 subunit (Alpha-adaptin A)

DE (Adaptor protein complex Ap-2 alpha-1 subunit) (Clathrin assembly

DE protein complex 2 alpha-A large chain) (100 kDa coated vesicle protein

DE A) (Plasma membrane adaptor HA2/AP2 adaptin alpha A subunit).

GN AP2A1 OR AD7AA OR CLAP1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI\_TaxID=10090;

RN [1]

RN SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE-89155572; PubMed-2564002;

RA Robinson M.S.;

RT "Cloning of cDNAs encoding two related 100-kD coated vesicle proteins

RL (alpha-adaptins).";

RL J. Cell Biol. 108:833-842(1989).

RN [2]

RN CHARACTERIZATION OF ISOFORMS A AND B.

RX PubMed=7593326;

RA Ball C.L., Hunt S.P., Robinson M.S.;

RT "Expression and localization of alpha-adaptin isoforms.";

RL J. Cell Sci. 108:2865-2875(1995).

CC -1- FUNCTION: Subunit of clathrin-associated adaptor protein complex 2

CC that plays a role in protein sorting in the late-Golgi/trans-Golgi

CC network (TGN) and/or endosomes. The AP complexes mediate both the

CC recruitment of clathrin to membranes and the recognition of

CC sorting signals within the cytosolic tails of transmembrane cargo

CC molecules. AP-2 complex seems to play a role in the recycling of

CC synaptic vesicle membranes from the presynaptic surface (by

CC similarity).

CC -1- SUBUNIT: Adaptor protein complex 2 (AP-2) is a heterotrimer

CC composed of two large adaptins (alpha1A/AP2A1 or alpha1B/AP2A1 or

CC alpha2/AP2A2 and beta1/AP2B1), a medium adaptin (mu2/AP2M1) and a

CC small adaptin (sigma2/AP2S1).

CC -1- SUBCELLULAR LOCATION: Component of the coat surrounding the

CC cytoplasmic face of coated vesicles located at the Golgi complex

CC (by similarity).

CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; A (shown here) and B; are

CC produced by alternative splicing.

CC -1- TISSUE SPECIFICITY: Isoform A is expressed only in neuronal tissue

CC and skeletal muscle. Isoform B is widely expressed.

CC -1- SIMILARITY: BELONGS TO THE ADAPTOR COMPLEXES LARGE SUBUNITS

CC FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL: X14971; CAA33096.1; -.

DR PIR: A30111; A30111.

DR MGD; MGI:101921; Ap2a1.

DR InterPro: IPR002553; Adaptin\_N.

DR InterPro: IPR003164; Alpha\_adaptin\_C.

DR InterPro: IPR001121; Gamma\_adaptin\_C.

DR Pfam: PF01602; Adaptin\_N; 1.

DR Pfam: PF02296; Alpha\_adaptin\_C; 1.

DR Pfam: PF02883; Alpha\_adaptin\_C2; 1.

DR Golgi stack; Protein transport; Transport; Coated pits; Endocytosis;

KW Alternative splicing.

FT VARSPLIC 706 727 MISSING (IN ISOFORM B).

FT SEQUENCE 977 AA; 107663 MW; F4ED87D3F9EF230A CRC64;

-----

Query Match 3.9%; Score 7; DB 1; Length 977;

Best Local Similarity 100.0%; Pred. No. 39;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SLTLKLP 20

Db 852 SLTLKLP 858

-----

RESULT 20

YC52\_XYLFA STANDARD; PRT; 1641 AA.

ID YC52\_XYLFA

```
AC Q9PDX7;
DT 15-JUN-2002 (Rel. 41, Created)
DE 15-JUN-2002 (Rel. 41, Last sequence update)
DE Hypothetical protein XF1252 precursor..
GN XF1252.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter R.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorfi H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kikajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A., Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A., Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA Vallada A., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Zeldanis J., Zetubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
CC -1- SIMILARITY: BELONGS TO THE UPF0192 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE003959; AAF84061.1;
CC DR EMBL; AE003959; AAF84061.1;
CC KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 31
FT CHAIN 32 1641 HYPOTHETICAL PROTEIN XF1252.
SQ SEQUENCE 1641 AA; 178920 MW; DF20B9B107364DB3 CRC64;
Query Match 3.9%; Score 7; DB 1; Length 1641;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 153 KGLLVE 159
DB 773 KGLLVE 779
RESULT 21
ZN32_HUMAN STANDARD; PRT; 52 AA.
AC P17041;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
```

```
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 32 (Zinc finger protein K0X30) (Fragment).
GN ZNF32 OR K0X30.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymphoid;
RX MEDLINE=91145339; PubMed=2288909;
RA Thiesen H.-J.;
RT "Multiple genes encoding zinc finger domains are expressed in human T
RL cells.";
RL New Biol. 2:363-374(1990).
CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X52361; CAA36587.1;
CC DR EMBL; X52361; CAA36587.1;
CC DR PIR; S10420; S10420.
CC DR HSP; P15822; IBBO.
CC DR Genew; HGNC:13095; ZNF32.
CC DR MIM; 194539;
CC DR InterPro; IPR000822; Znf_C2H2.
CC DR Pfam; PF000096; zf-C2H2; 2.
CC DR ProDom; PD000003; Znf_C2H2; 2.
CC DR SMART; SM00355; Znf_C2H2; 2.
CC DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
CC DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 2.
CC KW Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;
CC Nuclear protein; Repeat.
FT NON_TER 1 1
FT ZN_FING 1 23 C2H2-TYPE.
FT ZN_FING 29 51 C2H2-TYPE.
FT NON_TER 52 52
SQ SEQUENCE 52 AA; 6029 MW; 97E552B7C3918F78 CRC64;
Query Match 3.4%; Score 6; DB 1; Length 52;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 12 KGSLLT 17
DB 13 KGSLLT 18
RESULT 22
RS20_LACLA STANDARD; PRT; 77 AA.
AC Q9CEU5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S20.
GN RPST OR LL1739.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Boilotin A., Wincker P., Mauger S., Jaillon O., Malmarm K.,
```

```
RA Weissbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
CC -!- FUNCTION: Binds directly to 16S ribosomal RNA (By similarity).
CC -!- SIMILARITY: BELONGS TO THE S20P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE006404; AAK05837.1;
DR InterPro: IPR002583; Ribosomal_S20p.
DR Pfam: PF01649; Ribosomal_S20p; 1.
DR ProDom: PD004231; Ribosomal_S20p; 1.
DR Ribosomal protein; rRNA-binding; Complete proteome.
KW RIBOSOMAL PROTEIN; rRNA-BINDING; Complete proteome.
SQ SEQUENCE 77 AA; 8353 MW; 975F5D44BF145174 CRC64;

Query Match          3.4%; Score 6; DB 1; Length 77;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 DRAASK 103
DB 52 DRAASK 57

RESULT 23
RS20_MYCLE
ID RS20_MYCLE STANDARD; PRT; 86 AA.
AC Q3132;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S20.
GN RPST OR ML0604 OR MLC1536.08.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Elgmeler K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holtroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- FUNCTION: Binds directly to 16S ribosomal RNA (By similarity).
CC -!- SIMILARITY: BELONGS TO THE S20P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z99125; CAB16149.1;
DR EMBL: AL583919; CAC30112.1;
DR Leptoma: ML0604; -;
```

```
DR InterPro: IPR002583; Ribosomal_S20p.
DR Pfam: PF01649; Ribosomal_S20p; 1.
DR ProDom: PD004231; Ribosomal_S20p; 1.
DR Ribosomal protein; rRNA-binding; Complete proteome.
KW RIBOSOMAL PROTEIN; rRNA-BINDING; Complete proteome.
SQ SEQUENCE 86 AA; 9609 MW; 13B18F2730582790 CRC64;

Query Match          3.4%; Score 6; DB 1; Length 86;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 DRAASK 103
DB 59 DRAASK 64

RESULT 24
RS20_MYCTU
ID RS20_MYCTU STANDARD; PRT; 86 AA.
AC P71731;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S20.
GN RPST OR RV2412 OR MT2485 OR MTCY253.08C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeler K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Binds directly to 16S ribosomal RNA (By similarity).
CC -!- SIMILARITY: BELONGS TO THE S20P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z81368; CAB03746.1;
DR EMBL: AE007087; AAK46781.1;
DR TIGR: MT2485;
DR TubercuList: RV2412;
DR InterPro: IPR002583; Ribosomal_S20p.
DR Pfam: PF01649; Ribosomal_S20p; 1.
DR ProDom: PD004231; Ribosomal_S20p; 1.
DR Ribosomal protein; rRNA-binding; Complete proteome.
```

```
SQ SEQUENCE 86 AA; 9405 MW; 6B78BBFABC6DFIAC CRC64;
Query Match 3.4%; Score 6; DB 1; Length 86;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 DKAASK 103
| | | | |
Db 59 DKAASK 64

RESULT 25
RL35_PIG STANDARD; PRT; 94 AA.
AC Q29361;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 60S ribosomal protein L35 (fragment).
GN RPL35.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OC NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Small intestine;
RA Wintéroe A.K., Fredholm M., Davies W.;
RT "Evaluation and characterization of a porcine small intestine cDNA
RL Mamm. Genome 7:509-517(1996).
CC -1- SIMILARITY: BELONGS TO THE L29P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: F14766; CAA23237.1;
CC InterPro: IPR001854; Ribosomal_L29.
CC Pfam: PF00831; Ribosomal_L29; 1.
CC TIGRFAMs: TIGR00012; L29; 1.
CC PROSITE: PS00579; RIBOSOMAL_L29; 1.
CC Ribosomal protein.
CC NON_TER 1 1
CC NON_TER 94 94
SQ SEQUENCE 94 AA; 10926 MW; 19D8FEBD51F41D5F CRC64;
Query Match 3.4%; Score 6; DB 1; Length 94;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 LRGKKK 144
| | | | |
Db 7 LRGKKK 12

RESULT 26
Y004_BPHP1 STANDARD; PRT; 97 AA.
AC P51703;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 8.9 kDa protein in INT-C1 intergenic region (ORF4)
DE (ORF18).
OS Bacteriophage HP1.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC unclassified Myoviridae.
CC -----
OX NCBI_TaxID=10690;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HP1C1;
RX MEDLINE=95089704; PubMed=7997180;
RA Esposito D., Scocca J.J.;
RT "Identification of an HP1 phage protein required for site-specific
RL excision.";
RL Mol. Microbiol. 13:685-695(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HP1C1;
RX MEDLINE=96279738; PubMed=8710508;
RA Esposito D., Fitzmaurice W.P., Benjamin R.C., Goodman S.D.,
RA Waldman A.S., Scocca J.J.;
RT "The complete nucleotide sequence of bacteriophage HP1 DNA.";
RL Nucleic Acids Res. 24:2360-2368(1996).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: U24159; AAB09186.1;
CC DR EMBL; U24159; AAB09186.1;
CC KW Hypothetical protein.
SQ SEQUENCE 97 AA; 11266 MW; 7B76B495CEBDC139 CRC64;
Query Match 3.4%; Score 6; DB 1; Length 97;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 LTLKLP 20
| | | | |
Db 56 LTLKLP 61

RESULT 27
RK21_PORPU STANDARD; PRT; 104 AA.
AC P51209;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Chloroplast 50S ribosomal protein L21.
GN RPL21.
OS Porphyra purpurea.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX NCBI_TaxID=2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Avonport;
RA Reith M.E., Munnholland J.;
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RL genome.";
RL Plant Mol. Biol. Rep. 13:333-335(1995).
CC -1- FUNCTION: THIS PROTEIN BINDS TO 23S RIBOSOMAL RNA IN THE PRESENCE
CC OF PROTEIN L20 (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L21P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: U38804; AAC08095.1;
CC InterPro: IPR001787; Ribosomal_L21p.
```



```
DR Pfam: PF00829; Ribosomal_L21p; 1.
DR ProDom: PD003604; Ribosomal_L21p; 1.
DR TIGRFAMs: TIGR00061; L21; 1.
DR NCBI_TaxID=9606;
KW Ribosomal protein; rRNA-binding; Chloroplast.
SQ SEQUENCE 104 AA; 11877 MW; 76BEIC85F0DC355 CRC64;

Query Match      3.4%; Score 6; DB 1; Length 104;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 HLRGKK 143
DB 67 HLRGKK 72

RESULT 28
ID YHFI_ECOLI          STANDARD;          PRT; 120 AA.
AC P45551;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Hypothetical protein yhfY.
DE YHFI OR B3382.
GN YHFI OR B3382.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -1- SIMILARITY: TO H. INFLUENZAE HI0487.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U12465; AA51648.1; -.
CC EMBL: BC000348; AAH00348.1; -.
CC SWISS-2DPAGE: P42766; HUMAN.
CC Genew: HGNC:10344; RPL35.
CC InterPro: IPR001854; Ribosomal_L29.
CC Pfam: PF00831; Ribosomal_L29; 1.
CC TIGRFAMs: TIGR00012; L29; 1.
CC PROSITE: PS00579; RIBOSOMAL_L29; 1.
CC INIT MET 0
CC SEQUENCE 122 AA; 14420 MW; F44742BADAAC725D CRC64;
CC BY SIMILARITY.

Query Match      3.4%; Score 6; DB 1; Length 122;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 LRGKKK 144
DB 8 LRGKKK 13

RESULT 30
ID RL35_RAT            STANDARD;          PRT; 122 AA.
AC P17078;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE 60S ribosomal protein L35.
GN RPL35.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=90211326; PubMed=2322279;
RA Suzuki K., Olvera J., Wool I.G.;
RT "The primary structure of rat ribosomal protein L35.";
RL Biochem. Biophys. Res. Commun. 167:1377-1382(1990).
CC -1- SIMILARITY: BELONGS TO THE L29P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```

CC -----
DR EMBL; X51705; CAA36001.1; -.
DR PIR; S10390; R5RT35.
DR InterPro; IPR001854; Ribosomal_L29.
DR Pfam; PF00831; Ribosomal_L29; 1.
DR TIGRfams; TIGR00012; L29; 1.
DR PROSITE; PS00579; RIBOSOMAL_L29; 1.
KW Ribosomal protein.
FT INIT_MET 0
SQ SEQUENCE 122 AA; 14421 MW; 014742A17B1674EB CRC64;

Query Match 3.4%; Score 6; DB 1; Length 122;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 LRKKKK 144
DB 8 LRKKKK 13

RESULT 31
MSP1_GLORO
ID MSP1_GLORO STANDARD; PRT; 125 AA.
AC P53021;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Major sperm protein 1.
GN MSP-1.
OS Globodera rostochiensis (Golden nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.
OX NCBI_TaxID=31243;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ro1 / Mierenbos;
RA Novitski C.E., Brown S., Chen R., Corner A.S., Atkinson H.J.,
RA McPherson M.J.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC Sperm crawling. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD (BY
CC SIMILARITY).
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L24499; AAA29146.1; -.
CC HSSP; P27439; IMSP.
CC InterPro; IPR000535; MSP_domain.
CC Pfam; PF00635; MSP_domain; 1.
CC Cytokeleton; Acetylation; Sperm; Multigene family.
FT INIT_MET 0
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
SQ SEQUENCE 125 AA; 13858 MW; 0255C31F187549BC CRC64;

Query Match 3.4%; Score 6; DB 1; Length 125;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NGVLGP 8
DB 58 NGVLGP 63

RESULT 32
MSP3_GLORO
ID MSP3_GLORO STANDARD; PRT; 125 AA.
AC P53023;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Major sperm protein 3.
GN MSP-3.
OS Globodera rostochiensis (Golden nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.
OX NCBI_TaxID=31243;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ro1 / Mierenbos;
RA Novitski C.E., Brown S., Chen R., Corner A.S., Atkinson H.J.,
RA McPherson M.J.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC Sperm crawling. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD (BY
CC SIMILARITY).
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L24499; AAA29146.1; -.
CC HSSP; P27439; IMSP.
CC InterPro; IPR000535; MSP_domain.
CC Pfam; PF00635; MSP_domain; 1.
CC Cytokeleton; Acetylation; Sperm; Multigene family.
FT INIT_MET 0
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
SQ SEQUENCE 125 AA; 13858 MW; 0255C31F187549BC CRC64;

Query Match 3.4%; Score 6; DB 1; Length 125;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NGVLGP 8
DB 58 NGVLGP 63

RESULT 33
MSP3_GLORO
ID MSP3_GLORO STANDARD; PRT; 125 AA.
AC P53023;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Major sperm protein 3.
GN MSP-3.
OS Globodera rostochiensis (Golden nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.
OX NCBI_TaxID=31243;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ro1 / Mierenbos;
RA Novitski C.E., Brown S., Chen R., Corner A.S., Atkinson H.J.,
RA McPherson M.J.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC Sperm crawling. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD (BY

```

```

MSP2_GLORO
ID MSP2_GLORO STANDARD; PRT; 125 AA.
AC P53022;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Major sperm protein 2.
GN MSP-2.
OS Globodera rostochiensis (Golden nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.
OX NCBI_TaxID=31243;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ro1 / Mierenbos;
RA Novitski C.E., Brown S., Chen R., Corner A.S., Atkinson H.J.,
RA McPherson M.J.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC Sperm crawling. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD (BY
CC SIMILARITY).
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L24500; AAA29147.1; -.
CC HSSP; P27439; IMSP.
CC InterPro; IPR000535; MSP_domain.
CC Pfam; PF00635; MSP_domain; 1.
CC Cytokeleton; Acetylation; Sperm; Multigene family.
FT INIT_MET 0
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
SQ SEQUENCE 125 AA; 13801 MW; 1B04AD5756511B13 CRC64;

Query Match 3.4%; Score 6; DB 1; Length 125;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NGVLGP 8
DB 58 NGVLGP 63

RESULT 33
MSP3_GLORO
ID MSP3_GLORO STANDARD; PRT; 125 AA.
AC P53023;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Major sperm protein 3.
GN MSP-3.
OS Globodera rostochiensis (Golden nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.
OX NCBI_TaxID=31243;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ro1 / Mierenbos;
RA Novitski C.E., Brown S., Chen R., Corner A.S., Atkinson H.J.,
RA McPherson M.J.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC Sperm crawling. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD (BY

```

CC SIMILARITY).

CC -1- TISSUE SPECIFICITY: SPERM.

CC -1- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; L24501; AAA29148.1; .

DR HSSP; P27439; IMSP.

DR InterPro; IPR000535; MSP\_domain.

DR Pfam; PF00635; MSP\_domain; 1.

KW Cytoskeleton; Acetylation; Sperm; Multigene family.

FT INIT\_MET 0 0 BY SIMILARITY.

FT MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).

SO SEQUENCE 125 AA; 13844 MW; 0255C5A9C31E49BC CRC64;

Query Match 3.4%; Score 6; DB 1; Length 125;

Best Local Similarity 100.0%; Pred. No. 74;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NGVLGP 8

DB 58 NGVLGP 63

|||||

RESULT 34

FXSA\_SERMA STANDARD; PRT; 139 AA.

AC P37148;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE FXSA protein (suppressor of F exclusion of phase T7) (Fragment).

GN FXSA.

OS Serratia marcescens.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Serratia.

OX NCBI\_TaxID=615;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Sr41;

RX MEDLINE=95148727; PubMed=7846149;

RA Omori K., Akatsuka H., Komatsubara S.;

RT "Construction of a versatile promoter analysis vector and its use for

RL analysis of the Serratia marcescens aspartase promoter region.";

RL Plasmid 32:233-237(1994).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane

CC (Potential).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; D13252; BAA02517.1; .

DR PIR; PS0396; PS0396.

KW Transmembrane; Inner membrane.

FT TRANSMEM 17 37 POTENTIAL.

FT TRANSMEM 78 98 POTENTIAL.

FT NON\_TER 139 139

SO SEQUENCE 139 AA; 14996 MW; A897780614A5A09D CRC64;

Query Match 3.4%; Score 6; DB 1; Length 139;

Best Local Similarity 100.0%; Pred. No. 82;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SLTKL 19

DB 107 SLTKL 112

|||||

RESULT 35

MYG\_PSEGE STANDARD; PRT; 146 AA.

AC Q9DEP1;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Myoglobin.

GN MB.

OS Pseudochaenichthys georgianus (South Georgia icefish).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;

OC Notothenioidae; Channichthyidae; Pseudochaenichthys.

OX NCBI\_TaxID=52239;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Heart ventricle;

RX MEDLINE=97430285; PubMed=9284559;

RA Vayda M.E., Small D.J., Yuan M.-L., Costello L., Sidell B.D.;

RT "Conservation of the myoglobin gene among Antarctic notothenioid

RT fishes.";

RL Mol. Mar. Biol. Biotechnol. 6:207-216(1997).

CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES

CC THE MOVEMENT OF OXYGEN WITHIN MUSCLES.

CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U71055; AAG16643.1; .

DR HSSP; P02205; IMYT.

DR InterPro; IPR000971; Globin.

DR InterPro; IPR002335; Myoglobin.

DR Pfam; PF00042; globin; 2

DR PRINTS; PR00613; MYOGLOBIN.

DR PROSITE; PS01033; GLOBIN; 1.

KW Heme; Oxygen transport; Transport; Muscle.

FT INIT\_MET 0 0 BY SIMILARITY.

FT METAL 59 59 IRON (HEME DISTAL LIGAND).

FT METAL 88 88 IRON (HEME PROXIMAL LIGAND).

SO SEQUENCE 146 AA; 15574 MW; 8E006AD3EAE0FB40 CRC64;

Query Match 3.4%; Score 6; DB 1; Length 146;

Best Local Similarity 100.0%; Pred. No. 85;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 LVEADY 162

DB 12 LVEADY 17

|||||

RESULT 36

YXIK\_BACSU STANDARD; PRT; 153 AA.

ID YXIK\_BACSU

AC P42302;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypothetical protein yxik.

GN YXIK OR SS8B.

OS Bacillus subtilis.

```

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=97124196; PubMed=8969509;
RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
RA Miwa Y., Fujita Y.;
RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
RT containing the lic and cel loci, and creation of a 177 kb contig
RT covering the gnt-sacXY region.";
RL Microbiology 142:3113-3123(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azavedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.-K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konlingstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic C.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche B., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D83026; BAAL1690.1;
CC DR EMBL; 299124; CAB15950.1;
CC DR Subtilist; BG11140; ylik.
CC KW Hypothetical protein; Complete proteome.
CC SQ SEQUENCE 153 AA; 17649 MW; CFC1799C0BF2A4DC CRC64;

Query Match 3.4%; Score 6; DB 1; Length 153;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 93 AGLAYD 98
DB 39 AGLAYD 44
|||||

RESULT 37
GRPE_STRMU

```

```

ID GRPE_STRMU STANDARD; PRT; 180 AA.
AC O06941;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GRPE protein (HSP-70 cofactor).
GN GRPE.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=97426041; PubMed=9282745;
RA Jayaraman G.C., Penders J.E., Burne R.A.;
RT "Transcriptional analysis of the Streptococcus mutans hrca, grpe and
RT dnaK genes and regulation of expression in response to heat shock and
RT environmental acidification.";
RL Mol. Microbiol. 25:329-341(1997).
CC -!- FUNCTION: STIMULATES, JOINTLY WITH DNAJ, THE ATPASE ACTIVITY OF
CC DNAK. HELPS TO RELEASE ADP FROM DNAK THUS ALLOWING DNAK TO RECYCLE
CC MORE EFFICIENTLY (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE GRPE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U78296; AAC45611.1;
CC DR HSSP; P09372; 1DKG.
CC DR InterPro; IPS000740; Grpe.
CC DR Pfam; PF01025; Grpe; 1.
CC DR PROSITE; PS01071; GRPE; 1.
CC KW Chaperone; Heat shock.
CC SQ SEQUENCE 180 AA; 20621 MW; 5DD972412CDC5931 CRC64;

Query Match 3.4%; Score 6; DB 1; Length 180;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 67 KKEYK 72
DB 6 KKEYK 11
|||||

RESULT 38
ITC_ALOMA
ID ITC_ALOMA STANDARD; PRT; 184 AA.
AC P35812;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin/chymotrypsin inhibitor.
OS Alocasia macrorrhiza (Giant taro).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Araceae; Alocasia.
OX NCBI_TaxID=4456;
RN [1]
RP SEQUENCE.
RC STRAIN=CV. FUI-1;
RX MEDLINE=94191001; PubMed=8142459;
RA Argall M.E., Brandbury H.J., Shaw D.C.;
RT "Amino-acid sequence of a trypsin/chymotrypsin inhibitor from giant
RT taro (Alocasia macrorrhiza).";
RL Biochim. Biophys. Acta 1204:189-194(1994).
CC -!- FUNCTION: INHIBITS TRYPSIN AND ALPHA-CHYMOTRYPSIN.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SIMILARITY: BELONGS TO THE LEGUMINOUS KUNITZ-TYPE INHIBITOR

```

```
CC FAMILY.
DR PIR; S42753; S42753.
DR HSP; P07596; IAVA.
DR InterPro: IPR002160; Kunitz_legume.
DR Pfam: PF00197; Kunitz_legume; 1.
DR ProDom: PD000891; Kunitz_legume; 1.
DR SMART: SM00452; STI; 1.
DR PROSITE: PS00283; SOYBEAN_KUNITZ; FALSE_NEG.
KW Serine protease inhibitor.
FT DISULFID 39 84 BY SIMILARITY.
FT DISULFID 136 147 BY SIMILARITY.
FT ACT_SITE 56 57 REACTIVE BOND (POTENTIAL).
FT VARIANT 24 24 M -> A (IN 50% OF THE CHAINS).
FT VARIANT 50 50 E -> K (IN 25% OF THE CHAINS).
SQ SEQUENCE 184 AA; 15774 MW; 60536E9C15F472ED CRC64;

Query Match 3.4%; Score 6; DB 1; Length 184;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 NSRFAT 79
DB 113 NSRFAT 118

RESULT 39
PTH_BACHD STANDARD; PRT; 185 AA.
AC Q9KGJ3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Peptidyl-tRNA hydrolase (EC 3.1.1.29) (PTH).
GN PTH OR SPOVC OR BH0068.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.;
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- FUNCTION: THE NATURAL SUBSTRATE FOR THIS ENZYME MAY BE PEPTIDYL-
CC TRNAS WHICH DROP OFF THE RIBOSOME DURING PROTEIN SYNTHESIS (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: N-substituted aminoacyl-tRNA + H(2)O -> N-
CC substituted amino acid + tRNA.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PTH FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AP001507; BAB03787.1; .
CC HSP; P23932; 2PTH.
CC InterPro: IPR001328; Pept_tRNA_hydro.
CC Pfam: PF01195; Pept_tRNA_hydro; 1.
CC ProDom: PD005324; Pept_tRNA_hydro; 1.
CC TIGRFAMs: TIGR00447; pth; 1.
CC PROSITE: PS01195; PEPT_TRNA_HYDROL_1; 1.
CC PROSITE: PS01196; PEPT_TRNA_HYDROL_2; 1.
KW Hydrolase; Complete proteome.
```

```
SO SEQUENCE 185 AA; 20547 MW; AAC87F8B5354269D CRC64;

Query Match 3.4%; Score 6; DB 1; Length 185;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 63 EAFTKK 68
DB 170 EAFTKK 175

RESULT 40
DYS_HSVS7 STANDARD; PRT; 186 AA.
AC P27421;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Dihydrofolate reductase (EC 1.5.1.3) (DHFR).
GN 2 OR DHFR.
OS Herpesvirus saimiri (strain 484-77).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10382;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94269832; PubMed=8209420;
RA Geck P., Whitaker S.A., Medveczky M.M., Last T.J., Medveczky P.G.;
RT "Small RNA expression from the oncogenic region of a highly oncogenic
RT strain of herpesvirus saimiri.";
RL Virus Genes 8:23-34(1994).
CC -1- CATALYTIC ACTIVITY: 5,6,7,8-tetrahydrofolate + NADP(+) -> 7,8-
CC dihydrofolate + NADPH.
CC -1- PATHWAY: ESSENTIAL STEP FOR DE NOVO GLYCINE AND PURINE SYNTHESIS,
CC DNA PRECURSOR SYNTHESIS, AND FOR THE CONVERSION OF DUMP TO DTMP.
CC -1- SIMILARITY: BELONGS TO THE DIHYDROFOLATE REDUCTASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X58774; CAA41575.1; .
CC PIR; S14840; S14840.
CC HSP; P00374; IDHF.
CC InterPro: IPR001796; DHFR.
CC Pfam: PF00186; Dihfolate_red; 1.
CC PRINTS; PR00070; DHFR.
CC PROSITE; PS00075; DHFR; 1.
CC KW Oxidoreductase; NADP; One-carbon metabolism.
SQ SEQUENCE 186 AA; 21674 MW; 8249CD3C7ABE1BA2 CRC64;

Query Match 3.4%; Score 6; DB 1; Length 186;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 84 SLNDAL 89
DB 92 SLNDAL 97

Search completed: May 12, 2003, 09:47:58
Job time : 38 secs
```



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 09:44:40 ; Search time 31 Seconds  
(without alignments)  
1189.756 Million cell updates/sec

Title: US-09-995-493-52

Perfect score: 179

Sequence: 1 QHNGVLGPYIGKSLTKLP.....ADYTTKATANLYGLNLNRYR 179

Scoring table:

OLIGO Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organalle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	7.3	427	2	Q48051 haemophilus
2	13	7.3	449	2	Q9K2Q4 haemophilus
3	13	7.3	449	2	Q9KHF9 haemophilus
4	13	7.3	449	2	Q00754 haemophilus
5	13	7.3	450	2	Q9KHF4 haemophilus
6	13	7.3	450	2	Q9KHF3 haemophilus
7	13	7.3	451	2	Q9KHF2 haemophilus
8	13	7.3	451	2	Q9KHG0 haemophilus
9	13	7.3	451	2	Q9KHF8 haemophilus
10	13	7.3	451	2	Q9KHF7 haemophilus
11	13	7.3	451	2	Q9KHF6 haemophilus
12	13	7.3	454	2	Q9KHG1 haemophilus
13	13	7.3	455	2	Q9KHG5 haemophilus
14	13	7.3	455	2	Q9KHG4 haemophilus
15	13	7.3	455	2	Q9KHG3 haemophilus
16	13	7.3	455	2	Q9KHG2 haemophilus

90	7	3.9	556	4	Q96S68	Q96S68 homo sapien	163	6	3.4	142	2	Q9XAU1	Q9XAU1 pseudomonas
91	7	3.9	556	4	Q9BRD6	Q9BRD6 homo sapien	164	6	3.4	142	12	Q9YMS0	Q9YMS0 lymantria d
92	7	3.9	601	16	Q92T26	Q92T26 rhizobium m	165	6	3.4	143	5	Q9BK97	Q9BK97 axinella co
93	7	3.9	623	16	Q8XVB0	Q8XVB0 ralstonia s	166	6	3.4	143	17	Q9HST7	Q9HST7 halobacteri
94	7	3.9	671	16	Q92AQ0	Q92AQ0 listeria in	167	6	3.4	145	17	Q973Q2	Q973Q2 salifolobus
95	7	3.9	693	10	Q9W3E4	Q9W3E4 arabisdopsis	168	6	3.4	146	16	Q8ZS00	Q8ZS00 salmonella
96	7	3.9	731	10	Q93Y24	Q93Y24 arabisdopsis	169	6	3.4	146	17	Q9YAB9	Q9YAB9 aeropyrum p
97	7	3.9	734	10	Q80733	Q80733 arabisdopsis	170	6	3.4	147	5	Q61579	Q61579 drosophila
98	7	3.9	796	5	Q95YV4	Q95YV4 leishmania	171	6	3.4	147	13	Q9DGG0	Q9DGG0 sarda chili
99	7	3.9	849	3	Q12205	Q12205 saccharomyc	172	6	3.4	148	5	Q9NKU3	Q9NKU3 leishmania
100	7	3.9	885	16	Q92J25	Q92J25 rhizobium m	173	6	3.4	148	16	Q8ZMU0	Q8ZMU0 salmonella
101	7	3.9	904	10	Q9FUD9	Q9FUD9 phaseolus v	174	6	3.4	149	10	Q81057	Q81057 arabisdopsis
102	7	3.9	918	13	Q9W6U9	Q9W6U9 gallus gall	175	6	3.4	151	12	Q89374	Q89374 parametium
103	7	3.9	1094	4	Q9C0H5	Q9C0H5 homo sapien	176	6	3.4	151	16	Q9KTE2	Q9KTE2 vibrio chol
104	7	3.9	1330	16	Q8XRI4	Q8XRI4 ralstonia s	177	6	3.4	155	11	Q922V5	Q922V5 mus musculu
105	7	3.9	2060	3	Q9P980	Q9P980 pichia angu	178	6	3.4	155	16	Q8ZPV6	Q8ZPV6 salmonella
106	7	3.9	2340	11	Q64736	Q64736 mus musculu	179	6	3.4	156	2	Q8VRE5	Q8VRE5 hemophilus
107	7	3.9	2340	11	Q60705	Q60705 mus musculu	180	6	3.4	156	16	Q8Z6C0	Q8Z6C0 salmonella
108	7	3.9	3216	16	Q98Q29	Q98Q29 mycoplasma	181	6	3.4	157	5	Q9NB73	Q9NB73 heliothis z
109	7	3.9	4564	5	Q77075	Q77075 drosophila	182	6	3.4	158	5	Q26477	Q26477 schistocerc
110	6	3.4	30	6	Q29328	Q29328 sus scrofa	183	6	3.4	158	17	Q9YDG5	Q9YDG5 aeropyrum p
111	6	3.4	32	11	Q70125	Q70125 mus musculu	184	6	3.4	161	10	Q9SJB5	Q9SJB5 arabisdopsis
112	6	3.4	50	11	Q97640	Q97640 rattus norv	185	6	3.4	162	4	Q9Y4P9	Q9Y4P9 homo sapien
113	6	3.4	60	2	Q9R5H0	Q9R5H0 clostridium	186	6	3.4	163	11	Q9CW31	Q9CW31 mus musculu
114	6	3.4	55	10	Q9AQX7	Q9AQX7 g hypocheti	187	6	3.4	163	11	Q8R0W3	Q8R0W3 mus musculu
115	6	3.4	72	5	Q43983	Q43983 cryptospori	188	6	3.4	164	16	Q934T8	Q934T8 salmonella
116	6	3.4	75	2	Q9KV06	Q9KV06 vibrio chol	189	6	3.4	165	4	Q96PC6	Q96PC6 homo sapien
117	6	3.4	75	2	Q69271	Q69271 corynebacte	190	6	3.4	166	5	Q18389	Q18389 caenorhabdi
118	6	3.4	78	11	Q8RI02	Q8RI02 mus musculu	191	6	3.4	166	17	Q8UQJ7	Q8UQJ7 pyrococcus
119	6	3.4	78	16	Q99Z11	Q99Z11 streptococc	192	6	3.4	171	10	Q9FHW7	Q9FHW7 arabisdopsis
120	6	3.4	80	2	Q9WW17	Q9WW17 pseudomonas	193	6	3.4	171	10	Q94G68	Q94G68 triticum ae
121	6	3.4	80	2	Q9WW17	Q9WW17 pseudomonas	194	6	3.4	172	10	Q65283	Q65283 arabisdopsis
122	6	3.4	80	2	Q9WW16	Q9WW16 pseudomonas	195	6	3.4	173	13	Q91394	Q91394 clostrid co
123	6	3.4	80	2	Q9WVY0	Q9WVY0 pseudomonas	196	6	3.4	173	16	Q971H4	Q971H4 clostridium
124	6	3.4	80	2	Q9WVX9	Q9WVX9 pseudomonas	197	6	3.4	175	10	Q9M3X1	Q9M3X1 hordeum vul
125	6	3.4	80	2	Q9WVX4	Q9WVX4 pseudomonas	198	6	3.4	176	3	Q96V60	Q96V60 pneumocysti
126	6	3.4	80	2	Q9WW18	Q9WW18 pseudomonas	199	6	3.4	177	16	Q9RB78	Q9RB78 bacillus ha
127	6	3.4	81	17	Q97XS0	Q97XS0 sulfolobus	200	6	3.4	178	16	Q9RN73	Q9RN73 streptomyce
128	6	3.4	82	2	Q45911	Q45911 coxiella bu	201	6	3.4	179	2	Q44981	Q44981 borrelia af
129	6	3.4	82	6	Q95280	Q95280 sus scrofa	202	6	3.4	180	9	Q39995	Q39995 hordeum vul
130	6	3.4	84	16	Q9PE11	Q9PE11 xylella fas	203	6	3.4	180	9	Q9G0E8	Q9G0E8 lactococcus
131	6	3.4	86	10	Q9LMV6	Q9LMV6 arabisdopsis	204	6	3.4	180	17	Q8TZJ6	Q8TZJ6 pyrococcus
132	6	3.4	93	17	Q29367	Q29367 archaeoglob	205	6	3.4	181	2	Q93PA3	Q93PA3 microscilla
133	6	3.4	97	16	Q92PW1	Q92PW1 rhizobium m	206	6	3.4	181	11	Q9D7W1	Q9D7W1 mus musculu
134	6	3.4	99	9	Q8W6V7	Q8W6V7 cyanophage	207	6	3.4	181	16	Q927W6	Q927W6 chlamydia p
135	6	3.4	102	16	Q8R6M4	Q8R6M4 thermoanaer	208	6	3.4	184	2	Q34120	Q34120 borrelia bu
136	6	3.4	103	2	Q52873	Q52873 coxiella bu	209	6	3.4	184	5	Q95XG7	Q95XG7 caenorhabdi
137	6	3.4	105	10	Q49823	Q49823 citrus sine	210	6	3.4	184	10	Q9LWR4	Q9LWR4 oryza sativ
138	6	3.4	108	17	Q9Y915	Q9Y915 aeropyrum p	211	6	3.4	184	16	Q928Y9	Q928Y9 listeria in
139	6	3.4	113	12	Q8QXC1	Q8QXC1 tobacco lea	212	6	3.4	187	16	Q8Z1Z4	Q8Z1Z4 salmonella
140	6	3.4	116	4	Q43533	Q43533 homo sapien	213	6	3.4	187	16	Q8XB62	Q8XB62 escherichia
141	6	3.4	116	12	Q91R65	Q91R65 honeysuckle	214	6	3.4	188	16	Q9RWR2	Q9RWR2 deinococcus
142	6	3.4	117	10	Q9M3N4	Q9M3N4 hordeum vul	215	6	3.4	189	2	P94230	P94230 borrelia bu
143	6	3.4	118	12	Q9YPH2	Q9YPH2 broad bean	216	6	3.4	189	16	Q8Y2F1	Q8Y2F1 thermotoga
144	6	3.4	119	16	Q8ZF37	Q8ZF37 versinia pe	217	6	3.4	189	16	Q8Y2K0	Q8Y2K0 sulfolobus
145	6	3.4	120	2	Q9RDM8	Q9RDM8 lactobacill	218	6	3.4	189	17	Q97W64	Q97W64 sulfolobus
146	6	3.4	123	4	Q96QJ7	Q96QJ7 homo sapien	219	6	3.4	190	16	Q9RWB3	Q9RWB3 deinococcus
147	6	3.4	123	6	Q9BGU2	Q9BGU2 sus scrofa	220	6	3.4	190	17	Q29539	Q29539 archaeoglob
148	6	3.4	123	13	Q90YF7	Q90YF7 gallus gall	221	6	3.4	191	2	P94245	P94245 borrelia bu
149	6	3.4	123	13	Q90YF4	Q90YF4 ictalurus p	222	6	3.4	191	2	P94239	P94239 borrelia bu
150	6	3.4	124	2	Q54238	Q54238 rhizobium m	223	6	3.4	191	5	Q97064	Q97064 drosophila
151	6	3.4	124	4	Q96P30	Q96P30 homo sapien	224	6	3.4	191	10	Q9LVP6	Q9LVP6 arabisdopsis
152	6	3.4	124	16	Q92UX8	Q92UX8 rhizobium m	225	6	3.4	191	17	Q29160	Q29160 archaeoglob
153	6	3.4	124	16	Q92S03	Q92S03 rhizobium m	226	6	3.4	192	16	Q8ZMU1	Q8ZMU1 salmonella
154	6	3.4	124	17	Q58516	Q58516 pyrococcus	227	6	3.4	192	16	Q8Z346	Q8Z346 salmonella
155	6	3.4	124	17	Q9VA32	Q9VA32 aeropyrum p	228	6	3.4	192	16	Q8Z1F8	Q8Z1F8 salmonella
156	6	3.4	129	17	Q97XL6	Q97XL6 sulfolobus	229	6	3.4	194	2	P94247	P94247 borrelia bu
157	6	3.4	130	13	Q98903	Q98903 takifugu ru	230	6	3.4	195	2	Q54412	Q54412 streptomyce
158	6	3.4	133	2	Q9F7Y0	Q9F7Y0 pseudomonas	231	6	3.4	195	11	Q8VEL5	Q8VEL5 mus musculu
159	6	3.4	133	16	Q8X7N6	Q8X7N6 escherichia	232	6	3.4	195	16	Q9ZBR2	Q9ZBR2 streptomyce
160	6	3.4	138	11	Q9D3Q2	Q9D3Q2 mus musculu	233	6	3.4	196	16	Q9KEI9	Q9KEI9 bacillus ha
161	6	3.4	140	16	Q9HY44	Q9HY44 pseudomonas	234	6	3.4	200	8	Q9Z2C3	Q9Z2C3 phyllotis m
162	6	3.4	140	17	Q978X4	Q978X4 thermoplasm	235	6	3.4	200	10	Q9LWE5	Q9LWE5 oryza sativ



236	6	3.4	200	17	Q97B00	Q97b00 thermoplasm	309	6	3.4	247	2	Q00442	Q00442 saccharopol
237	6	3.4	201	2	P96514	P96514 borrelia ja	310	6	3.4	247	4	Q9H3H0	Q9h3h0 homo sapien
238	6	3.4	201	2	Q08234	Q08234 borrelia ta	311	6	3.4	247	17	Q9HQ44	Q9hqa4 halobacteri
239	6	3.4	201	2	P96516	P96516 borrelia ja	312	6	3.4	250	16	Q25124	Q25124 helicobacte
240	6	3.4	201	2	P96572	P96572 borrelia ja	313	6	3.4	250	16	Q92KB6	Q92kb6 helicobacte
241	6	3.4	201	2	P96573	P96573 borrelia ja	314	6	3.4	251	17	Q8TJY4	Q8tjy4 methanosarc
242	6	3.4	202	2	P96508	P96508 borrelia af	315	6	3.4	252	5	Q8T227	Q8t227 dictyostell
243	6	3.4	202	2	Q50626	Q50626 borrelia af	316	6	3.4	253	3	Q8WZQ8	Q8wzq8 neurospora
244	6	3.4	202	2	Q93097	Q93097 borrelia bu	317	6	3.4	253	5	Q9V5T6	Q9v5t6 drosophila
245	6	3.4	203	2	Q50619	Q50619 borrelia bu	318	6	3.4	253	16	Q50988	Q50988 borrelia bu
246	6	3.4	203	2	Q50624	Q50624 borrelia af	319	6	3.4	253	16	Q9HU06	Q9huu6 pseudomonas
247	6	3.4	203	2	Q44985	Q44985 borrelia ga	320	6	3.4	254	16	Q926L3	Q926l3 listeria in
248	6	3.4	203	10	Q82447	Q82447 nicotiana t	321	6	3.4	254	16	Q8Y108	Q8y108 brucella me
249	6	3.4	204	2	Q96509	Q96509 borrelia af	322	6	3.4	256	17	Q8TYA8	Q8tya8 methanopyru
250	6	3.4	204	2	Q44989	Q44989 borrelia af	323	6	3.4	257	5	O01336	O01336 caenorhabdi
251	6	3.4	204	16	Q9WYW1	Q9wyw1 thermotoga	324	6	3.4	258	10	Q940K8	Q940k8 arabidopsis
252	6	3.4	204	16	Q9A298	Q9a298 caulobacter	325	6	3.4	258	16	Q55434	Q55434 synechocyst
253	6	3.4	205	17	Q8TVN0	Q8tvn0 methanopyru	326	6	3.4	259	10	Q65544	Q65544 arabidopsis
254	6	3.4	206	16	Q9JQN8	Q9jqn8 neisseria m	327	6	3.4	260	16	Q92VL9	Q92vl9 rhizobium m
255	6	3.4	207	2	Q45175	Q45175 borrelia ga	328	6	3.4	261	2	Q8RP05	Q8rp05 azospirillu
256	6	3.4	207	10	Q96230	Q96230 allocasia ma	329	6	3.4	261	5	Q9V3C2	Q9v3c2 drosophila
257	6	3.4	208	16	Q8XCL3	Q8xcl3 escherichia	330	6	3.4	261	16	Q986Y9	Q986y9 rhizobium l
258	6	3.4	208	16	Q98OP6	Q98op6 mycoplasma	331	6	3.4	261	16	Q54172	Q54172 streptomyce
259	6	3.4	209	2	Q44671	Q44671 borrelia af	332	6	3.4	261	16	Q9WXQ6	Q9wxq6 thermotoga
260	6	3.4	209	2	Q49579	Q49579 borrelia ga	333	6	3.4	263	2	Q9RB25	Q9rb25 pectobacter
261	6	3.4	210	2	Q9EVT8	Q9evt8 thermus the	334	6	3.4	263	2	Q9RB24	Q9rb24 pectobacter
262	6	3.4	210	2	Q92679	Q92679 borrelia sp	335	6	3.4	263	2	Q9RB21	Q9rb21 pectobacter
263	6	3.4	210	2	Q9F435	Q9f435 klebsiella	336	6	3.4	263	2	Q9RB23	Q9rb23 pectobacter
264	6	3.4	210	5	Q9N6L5	Q9n6l5 leishmania	337	6	3.4	263	2	Q9XA48	Q9xta8 pectobacter
265	6	3.4	211	2	Q9S576	Q9s576 borrelia af	338	6	3.4	263	5	O16722	O16722 caenorhabdi
266	6	3.4	212	2	Q9KIM5	Q9kim5 borrelia af	339	6	3.4	263	16	Q8ZN27	Q8znz7 salmonella
267	6	3.4	212	10	Q9STQ4	Q9stq4 arabidopsis	340	6	3.4	263	16	Q8ZF19	Q8zfi9 versinia pe
268	6	3.4	213	17	Q8U2V0	Q8u2v0 pyrococcus	341	6	3.4	263	16	Q8Z667	Q8z667 salmonella
269	6	3.4	213	10	Q9FZ12	Q9fz12 solanum tub	342	6	3.4	263	16	Q8XCNA	Q8xcn4 escherichia
270	6	3.4	216	2	Q9RN47	Q9rn47 salmonella	343	6	3.4	264	16	Q8YF10	Q8yfl0 brucella me
271	6	3.4	217	3	Q12035	Q12035 saccharomyc	344	6	3.4	265	16	Q98B04	Q98bu4 rhizobium l
272	6	3.4	217	4	Q96BD2	Q96bd2 homo sapien	345	6	3.4	266	12	O72160	O72160 southern be
273	6	3.4	219	4	Q96PJ4	Q96pj4 homo sapien	346	6	3.4	267	2	O51673	O51673 paracoccus
274	6	3.4	219	4	Q9NTL0	Q9ntl0 homo sapien	347	6	3.4	267	16	Q97RN2	Q97rn2 streptococc
275	6	3.4	221	10	Q9FR12	Q9fr12 hemerocalli	348	6	3.4	267	16	Q8XY89	Q8xy89 ralstonia s
276	6	3.4	224	11	Q8RIW1	Q8riw1 mus musculu	349	6	3.4	268	10	Q90927	Q90927 lycopersico
277	6	3.4	225	4	Q96PJ1	Q96pj1 homo sapien	350	6	3.4	268	16	Q9A034	Q9a034 streptococc
278	6	3.4	226	4	Q96A05	Q96a05 homo sapien	351	6	3.4	268	16	Q97RA2	Q97ra2 streptococc
279	6	3.4	228	11	Q9CQC3	Q9cqc3 mus musculu	352	6	3.4	269	16	Q970T6	Q970t6 sulfolobus
280	6	3.4	228	16	Q25181	Q25181 helicobacte	353	6	3.4	270	3	O13758	O13758 schizosacch
281	6	3.4	228	16	Q25819	Q25819 helicobacte	354	6	3.4	270	4	Q96PJ2	Q96pj2 homo sapien
282	6	3.4	228	16	Q9XK06	Q9zk06 helicobacte	355	6	3.4	270	5	Q9VKA9	Q9vka9 drosophila
283	6	3.4	229	15	Q9HW77	Q9ht77 pseudomonas	356	6	3.4	270	17	Q9HRH9	Q9hrh9 halobacteri
284	6	3.4	230	5	O17452	O17452 drosophila	357	6	3.4	272	10	Q06765	Q06765 glycine max
285	6	3.4	231	17	Q8THH4	Q8thh4 methanosarc	358	6	3.4	272	16	Q989S2	Q989s2 rhizobium l
286	6	3.4	231	4	Q8WUI9	Q8wui9 homo sapien	359	6	3.4	273	4	Q92951	Q92951 homo sapien
287	6	3.4	231	16	Q9RJB6	Q9rjb6 streptomyce	360	6	3.4	273	10	Q9SS72	Q9ss72 arabidopsis
288	6	3.4	233	5	Q95YJ3	Q95yj3 samia cynth	361	6	3.4	273	10	Q8SB81	Q8sb81 oryza sativ
289	6	3.4	233	5	Q95YJ2	Q95yj2 samia cynth	362	6	3.4	273	13	O57659	O57659 gallus gall
290	6	3.4	233	16	Q985R3	Q985r3 rhizobium l	363	6	3.4	275	5	Q8SZ85	Q8sz85 drosophila
291	6	3.4	233	17	Q975G6	Q975g6 sulfolobus	364	6	3.4	277	17	Q9HQG4	Q9hqg4 halobacteri
292	6	3.4	233	17	Q96V77	Q96vt7 sulfolobus	365	6	3.4	278	10	Q948J9	Q948j9 arabidopsis
293	6	3.4	234	10	O04680	O04680 lycopersico	366	6	3.4	278	17	Q8TMV0	Q8tmv0 methanosarc
294	6	3.4	234	16	Q9RYI8	Q9ryi8 deinococcus	367	6	3.4	279	16	Q98RG8	Q98rg8 mycoplasma
295	6	3.4	236	5	Q9V157	Q9v157 drosophila	368	6	3.4	279	16	Q97KB0	Q97kb0 clostridium
296	6	3.4	236	10	Q40476	Q40476 nicotiana t	369	6	3.4	280	2	P96043	P96043 streptococc
297	6	3.4	236	16	Q8YQH0	Q8yqh0 anabaena sp	370	6	3.4	280	5	Q9VPB8	Q9vpb8 drosophila
298	6	3.4	236	16	O8XJG1	O8xjg1 clostridium	371	6	3.4	281	17	Q8ZSM1	Q8zsm1 pyroaculum
299	6	3.4	239	16	Q9A125	Q9a125 streptococc	372	6	3.4	281	2	Q9EY14	Q9ey14 streptomyce
300	6	3.4	239	16	Q9S233	Q9s233 streptomyce	373	6	3.4	281	10	Q8TBP2	Q8tbp2 homo sapien
301	6	3.4	240	2	Q9ADY8	Q9ady8 agrobacteri	374	6	3.4	281	10	Q9SN47	Q9sn47 arabidopsis
302	6	3.4	241	5	Q9N4P9	Q9n4p9 caenorhabdi	375	6	3.4	281	10	Q8S015	Q8s015 oryza sativ
303	6	3.4	241	17	Q8TWR2	Q8twr2 methanopyru	376	6	3.4	281	16	Q92BG8	Q92bg8 streptomyce
304	6	3.4	243	4	Q9H6F1	Q9h6f1 homo sapien	377	6	3.4	283	10	Q9FJ53	Q9fj53 arabidopsis
305	6	3.4	243	16	Q93J75	Q93j75 streptomyce	378	6	3.4	283	16	Q8X512	Q8x512 escherichia
306	6	3.4	244	16	O8YDQ9	O8ydg9 brucella me	379	6	3.4	283	16	Q8U8Y4	Q8u8y4 agrobacteri
307	6	3.4	245	16	Q9RV09	Q9rv09 deinococcus	380	6	3.4	283	16		
308	6	3.4	247	2	Q9JN61	Q9jn61 saccharopol	381	6	3.4	283	16		

382	6	3.4	285	16	Q9WXX3	Q9WXX3 thermotoga	455	6	3.4	307	16	Q8Y0X4	Q8Y0X4 ralstonia s
383	6	3.4	286	4	Q9P0C2	Q9P0C2 homo sapien	456	6	3.4	308	10	Q9ZU73	Q9ZU73 arabidopsis
384	6	3.4	286	16	Q8RE91	Q8RE91 fusobacteri	457	6	3.4	308	10	Q9M2P7	Q9M2P7 arabidopsis
385	6	3.4	288	5	O16767	O16767 caenorhabdi	458	6	3.4	308	11	O8VCS4	O8VCS4 mus musculu
386	6	3.4	288	16	O99ZJ3	O99ZJ3 streptococ	459	6	3.4	308	16	Q8XX55	Q8XX55 ralstonia s
387	6	3.4	289	2	Q93LZ1	Q93LZ1 rhizobium f	460	6	3.4	309	5	Q19681	Q19681 caenorhabdi
388	6	3.4	291	16	O8YLJ3	O8YLJ3 anabaena sp	461	6	3.4	309	17	Q976Y2	Q976Y2 sulfolobus
389	6	3.4	292	16	O9RJN6	O9RJN6 streptomyce	462	6	3.4	310	16	Q9X8R9	Q9X8R9 streptomyce
390	6	3.4	293	5	O8T923	O8T923 drosophila	463	6	3.4	312	16	Q92KJ9	Q92KJ9 rhizobium m
391	6	3.4	295	16	O9Z709	O9Z709 chlamydia p	464	6	3.4	313	10	O04828	O04828 hordeum vul
392	6	3.4	296	2	P72540	P72540 streptomyce	465	6	3.4	313	16	Q98EB4	Q98EB4 rhizobium l
393	6	3.4	296	10	Q9W336	Q9W336 arabidopsis	466	6	3.4	314	5	Q9N4I3	Q9N4I3 caenorhabdi
394	6	3.4	296	12	Q9IRK2	Q9IRK2 rabies viru	467	6	3.4	314	5	Q21202	Q21202 caenorhabdi
395	6	3.4	296	12	Q9IRJ1	Q9IRJ1 rabies viru	468	6	3.4	314	11	Q9Z2H7	Q9Z2H7 mus musculu
396	6	3.4	296	12	Q9IRJ0	Q9IRJ0 rabies viru	469	6	3.4	314	16	O8XJF2	O8XJF2 bruceella me
397	6	3.4	296	16	O984T0	O984T0 rhizobium l	470	6	3.4	315	4	O8TF65	O8TF65 homo sapien
398	6	3.4	297	12	O85414	O85414 rabies viru	471	6	3.4	317	2	O53126	Q53126 rhodococcus
399	6	3.4	297	12	Q9IRN9	Q9IRN9 rabies viru	472	6	3.4	317	2	P72260	P72260 rhodococcus
400	6	3.4	297	12	Q9IRN4	Q9IRN4 rabies viru	473	6	3.4	318	2	O8VXJ7	O8VXJ7 heliobacter
401	6	3.4	297	12	Q9IRN3	Q9IRN3 rabies viru	474	6	3.4	320	16	Q9ZBX6	Q9ZBX6 streptomyce
402	6	3.4	297	12	Q9IRN2	Q9IRN2 rabies viru	475	6	3.4	321	2	O59488	O59488 listeria iv
403	6	3.4	297	12	Q9IRN1	Q9IRN1 rabies viru	476	6	3.4	322	1	O50223	O50223 methanosarc
404	6	3.4	297	12	Q9IRN0	Q9IRN0 rabies viru	477	6	3.4	322	3	O74237	O74237 candida ten
405	6	3.4	297	12	Q9IRM9	Q9IRM9 rabies viru	478	6	3.4	322	4	O9BSM2	O9BSM2 homo sapien
406	6	3.4	297	12	Q9IRM8	Q9IRM8 rabies viru	479	6	3.4	323	5	Q95UR2	Q95UR2 tribolium c
407	6	3.4	297	12	Q9IRM7	Q9IRM7 rabies viru	480	6	3.4	323	10	Q9SB81	Q9SB81 arabidopsis
408	6	3.4	297	12	Q9IRM5	Q9IRM5 rabies viru	481	6	3.4	323	16	O8XXF7	O8XXF7 ralstonia s
409	6	3.4	297	12	Q9IRM4	Q9IRM4 rabies viru	482	6	3.4	323	17	O8TKF1	O8TKF1 methanosarc
410	6	3.4	297	12	Q9IRL9	Q9IRL9 rabies viru	483	6	3.4	324	17	O28240	O28240 archaeoglob
411	6	3.4	297	12	Q9IRL5	Q9IRL5 rabies viru	484	6	3.4	325	2	O46061	O46061 corynebacte
412	6	3.4	297	12	Q9IRL2	Q9IRL2 rabies viru	485	6	3.4	325	5	Q9NHA7	Q9NHA7 tribolium c
413	6	3.4	297	12	Q9IRK8	Q9IRK8 rabies viru	486	6	3.4	325	10	Q9S0U6	Q9S0U6 arabidopsis
414	6	3.4	297	12	Q9IRK4	Q9IRK4 rabies viru	487	6	3.4	326	16	Q8RDH3	Q8RDH3 thermoanaer
415	6	3.4	297	12	Q9IRK3	Q9IRK3 rabies viru	488	6	3.4	327	10	Q94D12	Q94D12 oryza sativ
416	6	3.4	297	12	Q9IRK0	Q9IRK0 rabies viru	489	6	3.4	327	10	O8RYG9	O8RYG9 oryza sativ
417	6	3.4	297	12	Q9IRJ9	Q9IRJ9 rabies viru	490	6	3.4	329	10	O9SR43	O9SR43 arabidopsis
418	6	3.4	297	12	Q9IRJ8	Q9IRJ8 rabies viru	491	6	3.4	330	2	Q9WXX6	Q9WXX6 burkholderi
419	6	3.4	297	12	Q9IRJ5	Q9IRJ5 rabies viru	492	6	3.4	330	10	Q42579	Q42579 arabidopsis
420	6	3.4	297	12	Q9IRI9	Q9IRI9 rabies viru	493	6	3.4	330	10	Q43730	Q43730 arabidopsis
421	6	3.4	297	12	Q9IRI8	Q9IRI8 rabies viru	494	6	3.4	331	16	O8Y0X8	O8Y0X8 ralstonia s
422	6	3.4	297	12	Q9IRI7	Q9IRI7 rabies viru	495	6	3.4	331	17	O980F4	O980F4 sulfolobus
423	6	3.4	297	12	Q9IRI6	Q9IRI6 rabies viru	496	6	3.4	332	1	O923Y8	Q923Y8 mus musculu
424	6	3.4	297	12	Q9IRI4	Q9IRI4 rabies viru	497	6	3.4	333	3	O94521	Q94521 schizosacch
425	6	3.4	297	12	Q9IRI2	Q9IRI2 rabies viru	498	6	3.4	333	16	O9A8C4	O9A8C4 caulobacter
426	6	3.4	297	12	Q9IRI0	Q9IRI0 rabies viru	499	6	3.4	334	5	Q93199	Q93199 caenorhabdi
427	6	3.4	297	12	Q9IRH9	Q9IRH9 rabies viru	500	6	3.4	335	5	O9N426	O9N426 caenorhabdi
428	6	3.4	297	12	Q9IRH8	Q9IRH8 rabies viru	501	6	3.4	336	10	O65453	O65453 arabidopsis
429	6	3.4	297	12	Q9I0X2	Q9I0X2 rabies viru	502	6	3.4	336	17	O9HPX4	O9HPX4 halobacteri
430	6	3.4	297	12	Q9I0M8	Q9I0M8 rabies viru	503	6	3.4	340	16	O8UJT5	O8UJT5 agrobacteri
431	6	3.4	297	12	Q9I0K5	Q9I0K5 rabies viru	504	6	3.4	341	3	Q96WN6	Q96WN6 trichoderma
432	6	3.4	297	12	Q9I0J6	Q9I0J6 rabies viru	505	6	3.4	345	2	Q9RBN7	Q9RBN7 rhodococcus
433	6	3.4	299	2	O68557	O68557 bacillus su	506	6	3.4	345	10	O9LYZ3	O9LYZ3 arabidopsis
434	6	3.4	299	2	Q9ADY4	Q9ADY4 agrobacteri	507	6	3.4	345	16	O97JD7	O97JD7 clostridium
435	6	3.4	299	16	Q92WP0	Q92WP0 rhizobium m	508	6	3.4	346	12	O9WYC3	O9WYC3 influenzavi
436	6	3.4	299	16	O8YPC0	O8YPC0 anabaena sp	509	6	3.4	347	5	O9GY39	O9GY39 leishmania
437	6	3.4	299	16	O8XT30	O8XT30 ralstonia s	510	6	3.4	347	10	O23221	O23221 arabidopsis
438	6	3.4	300	12	O8QL55	O8QL55 rubella vir	511	6	3.4	349	2	O31308	O31308 borrelia bu
439	6	3.4	300	16	O8RAW8	O8RAW8 thermoanaer	512	6	3.4	349	16	O51310	O51310 borrelia bu
440	6	3.4	301	5	O8SRD4	O8SRD4 encephalito	513	6	3.4	350	10	O96513	O96513 arabidopsis
441	6	3.4	301	16	O9KNL0	O9KNL0 vibrio chol	514	6	3.4	350	10	O9FX85	O9FX85 arabidopsis
442	6	3.4	301	16	O92NXL	O92NXL rhizobium m	515	6	3.4	350	10	O93WF9	O93WF9 oryza sativ
443	6	3.4	302	2	O87438	O87438 pseudomonas	516	6	3.4	350	11	O923H4	O923H4 mus musculu
444	6	3.4	302	2	O9API0	O9API0 legionella	517	6	3.4	352	2	O9F1Q7	O9F1Q7 acinetobact
445	6	3.4	302	16	Q9L5G0	Q9L5G0 salmonella	518	6	3.4	352	5	O02207	O02207 caenorhabdi
446	6	3.4	303	4	Q9BRQ3	Q9BRQ3 homo sapien	519	6	3.4	352	10	O9F8H5	O9F8H5 arabidopsis
447	6	3.4	303	10	O40953	O40953 porphyra pu	520	6	3.4	352	16	O9KU57	O9KU57 vibrio chol
448	6	3.4	304	16	O925B1	O925B1 streptomyce	521	6	3.4	354	13	O90WM7	O90WM7 fugu rubrip
449	6	3.4	304	16	Q925B1	Q925B1 streptomyce	522	6	3.4	356	16	O9PCN6	O9PCN6 xyella fas
450	6	3.4	305	4	O90UJ0	O90UJ0 homo sapien	523	6	3.4	356	16	O9KP76	O9KP76 vibrio chol
451	6	3.4	305	12	O36410	O36410 alcelaphine	524	6	3.4	357	5	O96507	O96507 crypanosoma
452	6	3.4	305	16	O8ZA96	O8ZA96 versinia pe	525	6	3.4	357	16	O9PPV8	O9PPV8 ureaplasma
453	6	3.4	306	2	O8VUX6	O8VUX6 staphylococ	526	6	3.4	358	17	O975E5	O975E5 sulfolobus
454	6	3.4	307	5	Q9BHX2	Q9BHX2 leishmania	527	6	3.4	358	17	Q970G1	Q970G1 sulfolobus

528	6	3.4	360	13	Q9YGC9	Q9ygc9 brachydanio	601	6	3.4	403	10	Q23109	O23109 arabidopsis
529	6	3.4	362	4	Q96S93	Q96s93 homo sapien	602	6	3.4	404	16	O85338	O85338 escherichia
530	6	3.4	362	5	Q22897	Q22897 caenorhabdi	603	6	3.4	405	4	O75295	O75295 homo sapien
531	6	3.4	362	16	Q8XVM7	Q8xvm7 ralstonia s	604	6	3.4	405	11	O9CZS7	O9czs7 mus musculu
532	6	3.4	363	2	O07652	O07652 cellvibrio	605	6	3.4	406	4	O9NMU3	O9nmw3 mus musculu
533	6	3.4	363	2	O52284	O52284 agrobacteri	606	6	3.4	406	6	O13010	O13010 homo sapien
534	6	3.4	363	16	O84198	O84198 chlamydia t	607	6	3.4	406	11	Q9R018	Q9r018 rattus norv
535	6	3.4	365	16	O9KFJ4	Q9kfj4 bacillus ha	608	6	3.4	407	16	Q9ZDJ2	Q9zdj2 rickettsia
536	6	3.4	365	16	O98P16	Q98p16 rhizobium l	609	6	3.4	407	16	O84558	O84558 chlamydia t
537	6	3.4	366	10	O9LVE7	Q9lve7 arabidopsis	610	6	3.4	411	16	O84558	O84558 chlamydia t
538	6	3.4	367	10	O9XHC9	Q9xhc9 hyacinthus	611	6	3.4	412	5	O965R2	O965r2 caenorhabdi
539	6	3.4	367	11	O9Q179	Q9q179 rattus norv	612	6	3.4	412	11	O9CU83	O9cu83 mus musculu
540	6	3.4	368	2	Q93QC6	Q93qc6 corynebacte	613	6	3.4	413	16	O8U9F9	O8uf99 agrobacteri
541	6	3.4	368	11	O8R1X0	Q8rlx0 mus musculu	614	6	3.4	414	10	O9C8X8	O9c8x8 arabidopsis
542	6	3.4	369	16	O9R2U0	Q9rzu0 deinococcus	615	6	3.4	415	10	O9SY27	O9sy27 arabidopsis
543	6	3.4	370	16	O9KVS3	Q9kvs3 vibrio chol	616	6	3.4	416	4	P78356	P78356 homo sapien
544	6	3.4	371	16	P95231	P95291 mycobacteri	617	6	3.4	416	11	O88377	O88377 rattus norv
545	6	3.4	374	5	O76950	O76950 helobdella	618	6	3.4	417	2	O9RPA9	O9rpa9 streptomyce
546	6	3.4	374	6	Q9TU81	Q9tu81 ovis aries	619	6	3.4	418	10	O9LKW7	O9lkw7 zea mays (m
547	6	3.4	374	10	Q8RUR1	Q8rur1 cryptomeria	620	6	3.4	418	10	O944N8	O944n8 arabidopsis
548	6	3.4	376	4	O8WXH3	Q8wxh3 homo sapien	621	6	3.4	419	13	O9Q8G1	O9q8g1 xenopus lae
549	6	3.4	377	10	O82563	O82563 anemia phyl	622	6	3.4	420	11	O88370	O88370 rattus norv
550	6	3.4	377	10	Q9M1P0	Q9mp0 arabidopsis	623	6	3.4	420	11	O81635	O81635 mus musculu
551	6	3.4	378	2	Q9EUS5	Q9eus5 streptomyce	624	6	3.4	420	11	O99M59	O99m59 mus musculu
552	6	3.4	379	10	Q9C9Q8	Q9c9q8 arabidopsis	625	6	3.4	420	16	O8XMD3	O8xmd3 clostridium
553	6	3.4	379	16	O9AK94	Q9ak94 streptomyce	626	6	3.4	421	4	O8TDX8	O8tdx8 homo sapien
554	6	3.4	381	16	O8Y316	O8y316 ralstonia s	627	6	3.4	421	11	O91XU3	O91xu3 mus musculu
555	6	3.4	382	16	O9YVU5	O9yv5 streptococc	628	6	3.4	421	11	O91XU3	O91xu3 mus musculu
556	6	3.4	383	2	Q9F5B6	Q9f5b6 agrobacteri	629	6	3.4	422	16	O8YK21	O8yk21 anabaena sp
557	6	3.4	383	10	Q9LEL8	Q9lel8 citrus sine	630	6	3.4	423	8	O9BBY7	O9by7 fasciola he
558	6	3.4	383	16	Q9CG58	Q9cg58 lactococcus	631	6	3.4	424	2	O50990	O50990 neisseria g
559	6	3.4	383	16	Q9K2X4	Q9k2x4 streptomyce	632	6	3.4	424	4	O9UF27	O9uf27 homo sapien
560	6	3.4	385	10	O9FNA7	Q9fna7 arabidopsis	633	6	3.4	424	16	O9K0C1	O9k0c1 neisseria m
561	6	3.4	385	10	O93X83	Q93x83 vitis labru	634	6	3.4	424	16	O9JVC6	O9jvc6 neisseria m
562	6	3.4	386	16	O97K26	Q97k26 clostridium	635	6	3.4	426	17	O972L9	O972l9 sulfobolus
563	6	3.4	386	16	O8ZHT6	Q8zht6 yersinia pe	636	6	3.4	427	10	O9CAX5	O9cax5 arabidopsis
564	6	3.4	386	16	O8YBH9	Q8ybh9 brucella me	637	6	3.4	428	10	O93WA3	O93wa3 oryza sativ
565	6	3.4	388	16	O9HW46	Q9hw46 pseudomonas	638	6	3.4	428	16	O9A548	O9a548 caulobacter
566	6	3.4	388	16	O9RY22	Q9ry22 deinococcus	639	6	3.4	429	16	O987R8	O987r8 rhizobium l
567	6	3.4	388	16	O92KC0	Q92kc0 rhizobium m	640	6	3.4	429	16	O8UEN8	O8uen8 agrobacteri
568	6	3.4	388	16	O8XV11	O8xv11 ralstonia s	641	6	3.4	429	16	O92R78	O92r78 rhizobium m
569	6	3.4	388	17	O8U056	O8u056 pyrococcus	642	6	3.4	430	16	O9KC09	O9kc09 bacillus ha
570	6	3.4	389	16	O98118	Q98118 rhizobium l	643	6	3.4	430	16	O9XAQ1	O9xaq1 streptomyce
571	6	3.4	389	16	O9RH89	Q9rh89 fusobacteri	644	6	3.4	431	5	O21891	O21891 caenorhabdi
572	6	3.4	390	2	Q9ZID8	Q9zid8 actinobacil	645	6	3.4	431	16	O95KG7	O95kg7 macaca fasc
573	6	3.4	390	13	P79753	P79753 fugu rubrip	646	6	3.4	431	16	O9PFB0	O9pfb0 xylella fas
574	6	3.4	391	12	O80871	O80871 non-a non-b	647	6	3.4	432	2	O9X592	O9x592 azospirillu
575	6	3.4	391	16	P72824	P72824 synechocyst	648	6	3.4	432	16	O9KT63	O9kt63 vibrio chol
576	6	3.4	391	16	O92721	Q92721 listeria in	649	6	3.4	432	16	O9AGJ9	O9agj9 caulobacter
577	6	3.4	391	16	O82RM5	O82rm5 salmonella	650	6	3.4	433	16	O9PPV2	O9ppv2 ureaplasma
578	6	3.4	391	16	O82986	O82986 salmonella	651	6	3.4	434	12	O9YVR1	O9yvr1 melanoplus
579	6	3.4	391	17	O972A7	O972a7 sulfobolus	652	6	3.4	434	16	O8RFR6	O8rf6 fusobacteri
580	6	3.4	393	4	O9BY52	O9by52 homo sapien	653	6	3.4	435	10	O94FR3	O94fr3 triticum ae
581	6	3.4	393	4	O9GSK0	Q9gsk0 homo sapien	654	6	3.4	437	5	O9BHB6	O9bhb6 plasmodium
582	6	3.4	393	4	O9BWS9	Q9bws9 homo sapien	655	6	3.4	437	5	O9BHB4	O9bhb4 plasmodium
583	6	3.4	394	2	O8RN55	Q8rn55 corynebacte	656	6	3.4	437	5	O9BH76	O9bh76 plasmodium
584	6	3.4	394	4	O9H836	Q9h836 homo sapien	657	6	3.4	437	5	O9BH63	O9bh63 plasmodium
585	6	3.4	394	5	O8SVG5	Q8svg5 encephalito	658	6	3.4	437	5	O9BH26	O9bh26 plasmodium
586	6	3.4	394	10	O9M0Y4	Q9m0y4 arabidopsis	659	6	3.4	437	5	O9BHS1	O9bhs1 plasmodium
587	6	3.4	396	5	O8SZ14	O8sz14 drosophila	660	6	3.4	437	5	O9BHS0	O9bhs0 plasmodium
588	6	3.4	396	16	O34367	O34367 bacillus su	661	6	3.4	437	5	O9BHS9	O9bhs9 plasmodium
589	6	3.4	397	16	O9RYF1	Q9ryf1 deinococcus	662	6	3.4	437	5	O9BHR8	O9bhr8 plasmodium
590	6	3.4	397	17	O8TLL1	O8tll1 methanosarc	663	6	3.4	437	5	O9BHR7	O9bhr7 plasmodium
591	6	3.4	398	11	O9D8A3	Q9d8a3 mus musculu	664	6	3.4	437	5	O9BHR6	O9bhr6 plasmodium
592	6	3.4	398	11	O8VCX9	Q8vcx9 mus musculu	665	6	3.4	437	5	O9BHR5	O9bhr5 plasmodium
593	6	3.4	400	16	O8ZPB5	O8zpb5 salmonella	666	6	3.4	437	5	O9BHR4	O9bhr4 plasmodium
594	6	3.4	400	16	O8Z763	O8z763 salmonella	667	6	3.4	437	5	O9BHR3	O9bhr3 plasmodium
595	6	3.4	400	16	O8Y091	Q8y091 ralstonia s	668	6	3.4	437	5	O9BHR2	O9bhr2 plasmodium
596	6	3.4	401	17	O9HNB5	Q9hnb5 halobacteri	669	6	3.4	437	5	O9BHR1	O9bhr1 plasmodium
597	6	3.4	401	17	O975F9	Q975f9 sulfobolus	670	6	3.4	437	5	O9BHR0	O9bhr0 plasmodium
598	6	3.4	402	5	O9BIM8	Q9bim8 plasmodium	671	6	3.4	437	5	O9BHQ9	O9bhq9 plasmodium
599	6	3.4	403	2	O9AE01	Q9ae01 amycolatops	672	6	3.4	437	5	O9BHQ8	O9bhq8 plasmodium
600	6	3.4	403	2	O9XDW7	Q9xdw7 rhodopseudo	673	6	3.4	437	5	O9BHQ7	O9bhq7 plasmodium

674	6	3.4	437	5	Q9BH06	Q9bhq6 plasmodium	747	6	3.4	469	5	Q9V3B8	Q9v3b8 drosophila
675	6	3.4	437	5	Q9BH05	Q9bhq5 plasmodium	748	6	3.4	469	10	Q9LVZ2	Q9lvz2 arbidopsi
676	6	3.4	437	5	Q9BH04	Q9bhq4 plasmodium	749	6	3.4	470	9	Q9FZV0	Q9fzv0 bacterioph
677	6	3.4	437	5	Q9BH03	Q9bhq3 plasmodium	750	6	3.4	471	16	Q9HWP9	Q9hwp9 pseudomonas
678	6	3.4	437	5	Q9BH02	Q9bhq2 plasmodium	751	6	3.4	471	17	Q8TXP4	Q8txp4 methanopyru
679	6	3.4	437	5	Q9BH01	Q9bhq1 plasmodium	752	6	3.4	475	16	Q98NA3	Q98na3 rhizobium l
680	6	3.4	437	5	Q9BH00	Q9bhq0 plasmodium	753	6	3.4	475	16	Q98PH4	Q98ph4 mycoplasma
681	6	3.4	437	5	Q9BHP9	Q9bhp9 plasmodium	754	6	3.4	476	6	Q9GMV3	Q9gmv3 macaca fasc
682	6	3.4	437	5	Q9BHP8	Q9bhp8 plasmodium	755	6	3.4	478	16	Q8UC05	Q8uc05 agrobacteri
683	6	3.4	437	5	Q9BHP7	Q9bhp7 plasmodium	756	6	3.4	480	5	Q17621	Q17621 caenorhabdi
684	6	3.4	437	5	Q9BHP6	Q9bhp6 plasmodium	757	6	3.4	480	5	Q9XVP8	Q9xvp8 caenorhabdi
685	6	3.4	437	5	Q9BHP5	Q9bhp5 plasmodium	758	6	3.4	481	5	Q27553	Q27553 cryptospori
686	6	3.4	437	5	Q9BHP4	Q9bhp4 plasmodium	759	6	3.4	481	16	Q98A95	Q98a95 rhizobium l
687	6	3.4	437	5	Q9BHP3	Q9bhp3 plasmodium	760	6	3.4	482	2	Q9AE73	Q9ae73 pseudomonas
688	6	3.4	437	5	Q9BHP2	Q9bhp2 plasmodium	761	6	3.4	483	4	Q8WVA1	Q8wva1 homo sapien
689	6	3.4	437	5	Q9BHP1	Q9bhp1 plasmodium	762	6	3.4	484	5	Q8SSH3	Q8ssh3 encephalito
690	6	3.4	437	5	Q9BHP0	Q9bhp0 plasmodium	763	6	3.4	484	10	Q65210	Q65210 nicotiana g
691	6	3.4	437	5	Q9BHN9	Q9bhn9 plasmodium	764	6	3.4	484	10	Q9SC82	Q9sc82 nicotiana t
692	6	3.4	437	5	Q9BHN8	Q9bhn8 plasmodium	765	6	3.4	484	17	Q8TN52	Q8tn52 methanosarc
693	6	3.4	437	5	Q9BHN7	Q9bhn7 plasmodium	766	6	3.4	485	2	Q9ZHU7	Q9zhu7 clostridium
694	6	3.4	437	5	Q9BHN6	Q9bhn6 plasmodium	767	6	3.4	486	3	Q01374	Q01374 neurospora
695	6	3.4	437	5	Q9BHN5	Q9bhn5 plasmodium	768	6	3.4	486	3	Q01376	Q01376 neurospora
696	6	3.4	437	5	Q9BHN4	Q9bhn4 plasmodium	769	6	3.4	487	16	P73225	P73225 synochocyst
697	6	3.4	437	5	Q9BHN3	Q9bhn3 plasmodium	770	6	3.4	488	10	Q8S133	Q8s133 oryza sativ
698	6	3.4	437	5	Q9BHN2	Q9bhn2 plasmodium	771	6	3.4	488	11	Q9DA59	Q9da59 mus musculu
699	6	3.4	437	10	Q9WE2	Q9we2 oryza sativ	772	6	3.4	488	13	Q91492	Q91492 torpedo cal
700	6	3.4	438	11	Q9JLB7	Q9jlb7 mus musculu	773	6	3.4	490	10	Q04270	Q04270 chlamydomon
701	6	3.4	439	5	Q9BLX8	Q9blx8 clona intes	774	6	3.4	491	16	Q97RS9	Q97rs9 streptococc
702	6	3.4	439	10	Q9FJV1	Q9fjv1 arbidopsi	775	6	3.4	492	5	Q97286	Q97286 plasmodium
703	6	3.4	439	17	Q8TN66	Q8tn66 methanosarc	776	6	3.4	494	3	Q8TFK2	Q8tfk2 aspergillus
704	6	3.4	442	16	Q92WF9	Q92wf9 rhizobium m	777	6	3.4	494	16	Q9CII7	Q9cii7 lactococcus
705	6	3.4	443	2	Q9L405	Q9l405 streptomyce	778	6	3.4	494	16	Q9RB68	Q9rb68 chlamydia p
706	6	3.4	444	2	Q93H31	Q93h31 streptomyce	779	6	3.4	497	5	Q76380	Q76380 oxytricha t
707	6	3.4	445	5	Q9U389	Q9u389 caenorhabdi	780	6	3.4	497	16	Q9A0V7	Q9aov7 streptococc
708	6	3.4	446	2	Q9F189	Q9f189 alcaligenes	781	6	3.4	498	16	Q8Y220	Q8y220 ralstonia s
709	6	3.4	446	10	P93632	P93632 zea mays (m	782	6	3.4	499	16	Q8ZA29	Q8za29 yersinia pe
710	6	3.4	446	16	Q98712	Q98712 rhizobium l	783	6	3.4	500	3	Q9UW95	Q9uw95 aspergillus
711	6	3.4	447	2	Q9AEN1	Q9aen1 bacillus ce	784	6	3.4	500	3	Q9UW94	Q9uw94 aspergillus
712	6	3.4	447	10	Q80885	Q80885 arbidopsi	785	6	3.4	500	3	Q9UW93	Q9uw93 aspergillus
713	6	3.4	447	16	Q9RUS4	Q9rus4 delnococcus	786	6	3.4	500	10	Q9SHK2	Q9shk2 arbidopsi
714	6	3.4	448	10	Q9SJR2	Q9sjr2 arbidopsi	787	6	3.4	502	16	Q9SFC3	Q9sfc3 staphylococ
715	6	3.4	448	16	Q97QY1	Q97qy1 streptococc	788	6	3.4	502	16	Q8ZPG3	Q8zpg3 salmonella
716	6	3.4	448	16	P96867	P96867 mycobacteri	789	6	3.4	505	11	Q8R2N2	Q8rzn2 mus musculu
717	6	3.4	449	10	Q9SK03	Q9sk03 arbidopsi	790	6	3.4	505	17	Q97VR4	Q97vr4 sulfolobus
718	6	3.4	450	5	Q9NDR2	Q9ndr2 drosophila	791	6	3.4	506	16	Q9XBN5	Q9xbn5 streptococc
719	6	3.4	450	5	Q9V3H0	Q9v3h0 drosophila	792	6	3.4	509	2	Q9X4G1	Q9x4g1 escherichia
720	6	3.4	450	16	P71758	P71758 mycobacteri	793	6	3.4	509	2	Q9LAR4	Q9lar4 bradyrhizob
721	6	3.4	453	2	Q51971	Q51971 pseudomonas	794	6	3.4	509	2	Q06958	Q06958 salmonella
722	6	3.4	453	2	Q9R928	Q9r928 pseudomonas	795	6	3.4	509	16	Q8R8W1	Q8r8w1 thermomanaer
723	6	3.4	453	5	Q8T887	Q8t887 clona intes	796	6	3.4	510	11	Q9JLB8	Q9jlb8 mus musculu
724	6	3.4	453	10	Q9MAG4	Q9mag4 arbidopsi	797	6	3.4	512	5	Q17559	Q17559 caenorhabdi
725	6	3.4	453	16	Q8XHP9	Q8xhp9 clostridium	798	6	3.4	512	11	Q925H0	Q925h0 mus musculu
726	6	3.4	454	2	Q9L4D7	Q9l4d7 xanthomonas	799	6	3.4	514	10	Q9ZS56	Q9zs56 arbidopsi
727	6	3.4	454	5	Q18697	Q18697 caenorhabdi	800	6	3.4	517	8	Q9M7A7	Q9mia7 typhlonecte
728	6	3.4	454	10	Q9M3V7	Q9m3v7 asparagus o	801	6	3.4	517	10	Q8S0S0	Q8s0s0 oryza sativ
729	6	3.4	454	16	Q97I25	Q97i25 clostridium	802	6	3.4	518	16	Q9CF65	Q9cf65 lactococcus
730	6	3.4	455	6	Q9N134	Q9n134 cryotcolagus	803	6	3.4	519	6	Q95JS3	Q95js3 macaca fasc
731	6	3.4	456	13	Q918S4	Q9i8s4 xenopus lae	804	6	3.4	519	10	Q9XHD4	Q9xhd4 petunia hyb
732	6	3.4	457	2	Q9AL89	Q9al89 agrobacteri	805	6	3.4	520	3	Q12610	Q12610 emeritella
733	6	3.4	458	10	Q9FHR4	Q9fhr4 arbidopsi	806	6	3.4	521	2	Q9RWF9	Q9rmf9 streptomyce
734	6	3.4	460	17	Q8T1J0	Q8t1j0 methanosarc	807	6	3.4	523	4	Q8TDC3	Q8tdg3 homo sapien
735	6	3.4	462	16	Q9KES0	Q9kes0 bacillus ha	808	6	3.4	525	5	Q9N9F1	Q9n9f1 plasmodium
736	6	3.4	463	12	Q85291	Q85291 potato mop-	809	6	3.4	525	10	Q48696	Q48696 arbidopsi
737	6	3.4	463	12	Q91V54	Q91v54 potato mop-	810	6	3.4	526	4	Q9H703	Q9h703 homo sapien
738	6	3.4	464	3	Q93914	Q93914 aspergillus	811	6	3.4	526	5	Q9NG01	Q9ng01 plasmodium
739	6	3.4	465	2	Q65943	Q65943 pseudomonas	812	6	3.4	526	5	Q9N9C0	Q9n9c0 plasmodium
740	6	3.4	465	2	Q8VM12	Q8vm12 pseudomonas	813	6	3.4	526	5	Q9N9F9	Q9n9f9 plasmodium
741	6	3.4	465	4	Q9NVZ6	Q9nvz6 homo sapien	814	6	3.4	526	5	Q9N9F8	Q9n9f8 plasmodium
742	6	3.4	466	12	Q8QKZ1	Q8qkz1 turnip mosa	815	6	3.4	526	5	Q9N9F7	Q9n9f7 plasmodium
743	6	3.4	467	16	Q9S2Y4	Q9s2y4 streptomyce	816	6	3.4	526	5	Q9N9F6	Q9n9f6 plasmodium
744	6	3.4	467	17	Q96VH4	Q96vh4 sulfolobus	817	6	3.4	526	5	Q9N9F5	Q9n9f5 plasmodium
745	6	3.4	468	4	Q961I7	Q961i7 homo sapien	818	6	3.4	526	5	Q9N9F4	Q9n9f4 plasmodium
746	6	3.4	469	1	Q31075	Q31075 methanosarc	819	6	3.4	526	5	Q9N9F3	Q9n9f3 plasmodium

820	6	3.4	526	5	Q9N9F2	Q9n9f2 plasmodium	893	6	3.4	575	4	Q96E88	Q96e88 homo sapien
821	6	3.4	526	5	Q9N9F0	Q9n9f0 plasmodium	894	6	3.4	575	4	Q8TDW4	Q8tdw4 homo sapien
822	6	3.4	526	5	Q9N9E9	Q9n9e9 plasmodium	895	6	3.4	575	16	Q98K94	Q98k94 rhizobium l
823	6	3.4	526	5	Q9N9E8	Q9n9e8 plasmodium	896	6	3.4	576	16	Q9FC77	Q9fc77 streptomyce
824	6	3.4	526	5	Q9N9E7	Q9n9e7 plasmodium	897	6	3.4	577	10	Q96RU0	Q96ru0 homo sapien
825	6	3.4	526	5	Q9N9E6	Q9n9e6 plasmodium	898	6	3.4	577	10	Q9XEW1	Q9xew1 arabidopsis
826	6	3.4	526	5	Q9N9E5	Q9n9e5 plasmodium	899	6	3.4	577	16	Q913W8	Q913w8 pseudomonas
827	6	3.4	526	5	Q9N9E4	Q9n9e4 plasmodium	900	6	3.4	579	10	Q9FG46	Q9fg46 arabidopsis
828	6	3.4	526	5	Q9N9E3	Q9n9e3 plasmodium	901	6	3.4	579	10	Q8QZS3	Q8qzs3 mus musculus
829	6	3.4	526	5	Q9N9E2	Q9n9e2 plasmodium	902	6	3.4	580	5	Q95Y24	Q95y24 caenorhabdi
830	6	3.4	526	5	Q9N9E1	Q9n9e1 plasmodium	903	6	3.4	580	16	Q8Y2A4	Q8y2a4 ralsstonia s
831	6	3.4	526	5	Q9N9E0	Q9n9e0 plasmodium	904	6	3.4	580	16	Q8Y199	Q8y199 ralsstonia s
832	6	3.4	526	5	Q9N9D9	Q9n9d9 plasmodium	905	6	3.4	581	17	Q8ZYM0	Q8zym0 pyrobaculum
833	6	3.4	526	10	Q9LJS6	Q9ljse arabidopsis	906	6	3.4	582	2	Q9X337	Q9x337 bacillus an
834	6	3.4	529	16	Q92XG5	Q92xg5 rhizobium m	907	6	3.4	582	16	Q9PJU5	Q9pju5 chlamydia m
835	6	3.4	529	16	Q8UJH1	Q8uji1 agrobacteri	908	6	3.4	583	10	Q9ZV15	Q9zvl5 arabidopsis
836	6	3.4	531	10	Q8FV34	Q8fv34 picea abies	909	6	3.4	583	10	Q9FU94	Q9fu94 oryza sativ
837	6	3.4	532	2	Q85154	Q85154 photorhabdi	910	6	3.4	584	16	Q84453	Q84453 chlamydia t
838	6	3.4	532	12	Q36160	Q36160 vibrio chol	911	6	3.4	588	2	Q54456	Q54456 serratia ma
839	6	3.4	532	16	Q912X5	Q912x5 pseudomonas	912	6	3.4	588	2	Q9XB64	Q9xb64 erwinita amy
840	6	3.4	533	3	Q43013	Q43013 schizosacch	913	6	3.4	588	13	Q9PU57	Q9pu57 brachydanio
841	6	3.4	533	10	Q9AX14	Q9ax14 plisum sativ	914	6	3.4	589	11	Q91WQ5	Q91wq5 mus musculus
842	6	3.4	533	16	Q8U6S7	Q8u6s7 agrobacteri	915	6	3.4	590	4	Q9H2P2	Q9h2p2 homo sapien
843	6	3.4	534	10	Q94F71	Q94f71 triticum ae	916	6	3.4	590	5	Q9N9R4	Q9n9r4 leishmania
844	6	3.4	535	16	Q9A575	Q9a575 caulobacter	917	6	3.4	591	2	Q8VUH3	Q8vuh3 streptococ
845	6	3.4	537	16	Q99XX7	Q99xx7 streptococ	918	6	3.4	591	10	Q9LV08	Q9lv08 arabidopsis
846	6	3.4	539	10	Q94AD8	Q94ad8 arabidopsis	919	6	3.4	592	5	Q25745	Q25745 plasmodium
847	6	3.4	540	2	Q8RJN5	Q8rjn5 xanthomonas	920	6	3.4	592	5	Q25749	Q25749 plasmodium
848	6	3.4	540	5	Q8SSC9	Q8ssc9 encephalito	921	6	3.4	593	13	Q9PUC2	Q9puc2 brachydanio
849	6	3.4	544	4	Q8TDW2	Q8tdw2 homo sapien	922	6	3.4	593	16	Q8ZNE5	Q8zne5 salmonella
850	6	3.4	544	10	Q9C509	Q9c509 arabidopsis	923	6	3.4	594	4	Q9H9X0	Q9h9x0 homo sapien
851	6	3.4	544	16	Q9A195	Q9a195 streptococ	924	6	3.4	594	10	Q9SMQ2	Q9smq2 arabidopsis
852	6	3.4	545	17	Q8THY6	Q8thy6 methanosarc	925	6	3.4	596	2	Q9RBJ7	Q9rbj7 xanthomonas
853	6	3.4	546	16	Q92T33	Q92t33 rhizobium m	926	6	3.4	596	4	Q8TE89	Q8te89 homo sapien
854	6	3.4	547	2	Q44170	Q44170 actinobacil	927	6	3.4	596	10	Q941Y1	Q941y1 oryza sativ
855	6	3.4	547	2	Q44124	Q44124 actinobacil	928	6	3.4	599	10	Q94BN1	Q94bn1 arabidopsis
856	6	3.4	547	2	Q44168	Q44168 actinobacil	929	6	3.4	599	10	Q8W561	Q8w561 nicotiana t
857	6	3.4	548	10	Q9C6L8	Q9c6l8 arabidopsis	930	6	3.4	601	13	Q42500	Q42500 meleagris g
858	6	3.4	549	4	Q9NCS3	Q9ncs3 homo sapien	931	6	3.4	604	5	Q25746	Q25746 plasmodium
859	6	3.4	549	11	Q9D006	Q9d006 mus musculus	932	6	3.4	604	5	Q25747	Q25747 plasmodium
860	6	3.4	549	11	Q9JLB9	Q9jlb9 mus musculus	933	6	3.4	605	5	Q9NFW9	Q9nfw9 plasmodium
861	6	3.4	550	16	Q8U6K0	Q8u6k0 agrobacteri	934	6	3.4	605	5	Q9GVB7	Q9gvb7 plasmodium
862	6	3.4	551	5	Q21768	Q21768 caenorhabdi	935	6	3.4	606	4	Q9BZ20	Q9bz20 homo sapien
863	6	3.4	553	11	Q9CVV4	Q9cyy4 mus musculus	936	6	3.4	610	10	Q9S791	Q9s791 arabidopsis
864	6	3.4	555	4	Q8TDW1	Q8tdw1 homo sapien	937	6	3.4	610	16	Q8RDR5	Q8dr5 fusobacteri
865	6	3.4	556	10	Q9SDI9	Q9sdi9 oryza sativ	938	6	3.4	611	16	Q9FC95	Q9fc95 streptomyce
866	6	3.4	557	1	Q59662	Q59662 pyrodictum	939	6	3.4	613	5	Q8SXN5	Q8sxn5 drosophila
867	6	3.4	557	10	Q93VF8	Q93vf8 oryza sativ	940	6	3.4	614	4	Q96FE5	Q96fe5 homo sapien
868	6	3.4	557	10	Q8RWK1	Q8rwk1 arabidopsis	941	6	3.4	614	6	Q9N008	Q9n008 macaca fasc
869	6	3.4	557	16	Q9I406	Q9i406 pseudomonas	942	6	3.4	614	11	Q9D1T0	Q9d1t0 mus musculus
870	6	3.4	558	4	Q8TDW3	Q8tdw3 homo sapien	943	6	3.4	615	12	Q91LC9	Q91lc9 white spot
871	6	3.4	558	10	Q9FGE3	Q9fge3 arabidopsis	944	6	3.4	615	16	Q98Q57	Q98q57 mycoplasma
872	6	3.4	558	16	Q9WYH2	Q9wyh2 thermotoga	945	6	3.4	617	16	Q8RMW9	Q8rm9 thermomane
873	6	3.4	558	16	Q8YCC4	Q8ycg4 brucella me	946	6	3.4	619	2	Q69221	Q69221 azotobacter
874	6	3.4	559	4	Q9H796	Q9h796 homo sapien	947	6	3.4	620	5	Q00784	Q00784 plasmodium
875	6	3.4	559	5	Q9V703	Q9v703 drosophila	948	6	3.4	620	16	Q007239	Q007239 mycobacteri
876	6	3.4	561	4	Q9BXF0	Q9bxfo homo sapien	949	6	3.4	621	4	Q98UL6	Q9bul6 homo sapien
877	6	3.4	561	10	Q82601	Q82601 arabidopsis	950	6	3.4	621	4	Q96G02	Q96g02 homo sapien
878	6	3.4	562	12	Q93076	Q93076 marine birt	951	6	3.4	621	5	Q9TY48	Q9ty48 plasmodium
879	6	3.4	562	16	Q8YUVO	Q8yuv0 anabaena sp	952	6	3.4	621	16	Q912T8	Q912t8 pseudomonas
880	6	3.4	563	10	Q945K2	Q945k2 prunus dulc	953	6	3.4	622	5	Q9GZ05	Q9gz05 plasmodium
881	6	3.4	569	10	Q8VZF1	Q8vzf1 arabidopsis	954	6	3.4	622	5	Q96TK3	Q96tk3 plasmodium
882	6	3.4	570	2	Q8RS16	Q8rsi6 uncultured	955	6	3.4	622	5	Q967K2	Q967k2 plasmodium
883	6	3.4	570	17	Q97WS9	Q97ws9 sulfolobus	956	6	3.4	622	5	Q967K1	Q967k1 plasmodium
884	6	3.4	571	10	Q9LS00	Q9lsq0 arabidopsis	957	6	3.4	622	5	Q967K0	Q967k0 plasmodium
885	6	3.4	571	10	Q9SM48	Q8sw48 arabidopsis	958	6	3.4	622	5	Q967J9	Q967j9 plasmodium
886	6	3.4	571	16	Q9CMT4	Q9cmt4 pasteurella	959	6	3.4	622	5	Q967J8	Q967j8 plasmodium
887	6	3.4	572	2	Q9ARG8	Q9arg8 comamonas t	960	6	3.4	622	5	Q967J7	Q967j7 plasmodium
888	6	3.4	572	2	P71187	P71187 enterobacte	961	6	3.4	622	5	Q967J6	Q967j6 plasmodium
889	6	3.4	572	2	Q936Y6	Q936y6 pseudomonas	962	6	3.4	622	5	Q967J5	Q967j5 plasmodium
890	6	3.4	572	11	Q921N6	Q921n6 mus musculus	963	6	3.4	622	5	Q967J4	Q967j4 plasmodium
891	6	3.4	572	16	Q9RX48	Q9rx48 deinococcus	964	6	3.4	622	5	Q967J3	Q967j3 plasmodium
892	6	3.4	574	16	P96591	P96591 bacillus su	965	6	3.4	622	5	Q967J2	Q967j2 plasmodium

```

966 6 3.4 622 5 Q967J1 plasmodium
967 6 3.4 622 5 Q967J0 plasmodium
968 6 3.4 622 5 Q967I9 plasmodium
969 6 3.4 622 5 Q967I8 plasmodium
970 6 3.4 622 5 Q967I7 plasmodium
971 6 3.4 622 5 Q967I6 plasmodium
972 6 3.4 622 5 Q967I5 plasmodium
973 6 3.4 622 5 Q967I4 plasmodium
974 6 3.4 622 5 Q967I3 plasmodium
975 6 3.4 622 5 Q967I2 plasmodium
976 6 3.4 622 5 Q95N21 plasmodium
977 6 3.4 622 5 Q95N04 plasmodium
978 6 3.4 622 5 Q95N02 plasmodium
979 6 3.4 622 5 Q94661 plasmodium
980 6 3.4 626 4 Q9H967 homo sapien
981 6 3.4 627 10 Q9SL48
982 6 3.4 629 4 Q95671
983 6 3.4 630 10 Q94GH1
984 6 3.4 631 5 Q45873
985 6 3.4 633 10 Q9SI73
986 6 3.4 633 10 Q9FV33
987 6 3.4 633 17 Q9Y176
988 6 3.4 633 17 Q9HP10
989 6 3.4 634 5 Q22585
990 6 3.4 636 2 Q938A9
991 6 3.4 642 4 Q9P263
992 6 3.4 645 8 Q63620
993 6 3.4 645 16 Q9PNH8
994 6 3.4 645 16 Q9RDB4
995 6 3.4 645 16 Q9PBT3
996 6 3.4 650 2 Q8RR67
997 6 3.4 653 10 Q9SS77
998 6 3.4 653 16 Q9I212
999 6 3.4 657 10 Q9LFX6
1000 6 3.4 658 4 Q9BQ26 homo sapien

```

## ALIGNMENTS

```

RESULT 1
Q48051 ID Q48051 PRELIMINARY; PRT; 427 AA.
AC Q48051;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Outer membrane protein P1 (Fragment).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F3029;
RA Read R.B., Frost J.B., Kort K., Myers S.D., Lesse A.J.;
RT "DNA sequence analysis and restriction fragment length polymorphisms
RT of the P1 gene of Haemophilus influenzae biogroup aegyptius associated
RT with Brazilian purpuric fever.";
RL Infect. Immun. 64:3665-3672(1996).
DR EMBL; U28808; AAC44459.1; -
FT NON_TER 1 1
FT VARIANT 61 61 D -> Y.
FT VARIANT 265 265 D -> N.
FT VARIANT 385 385 Y -> H.
SQ SEQUENCE 427 AA; 46489 MW; 80620BA6FA0DD27 CRC64;

```

Query Match 7.3%; Score 13; DB 2; Length 427;  
 Best Local Similarity 100.0%; Pred. No. 6.5e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133

Db 370 GATYKFTPNLSVD 382

## RESULT 2

```

Q9K2Q4 ID Q9K2Q4 PRELIMINARY; PRT; 449 AA.
AC Q9K2Q4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Outer membrane protein P1.
GN OMPP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC9833, AND 6255;
MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL; AF260369; AAF97583.1; -
DR EMBL; AF260368; AAF97582.1; -
SQ SEQUENCE 449 AA; 48518 MW; BBDB55BB2064E073 CRC64;

```

Query Match 7.3%; Score 13; DB 2; Length 449;  
 Best Local Similarity 100.0%; Pred. No. 6.8e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133

Db 390 GATYKFTPNLSVD 402

## RESULT 3

```

Q9KHF9 ID Q9KHF9 PRELIMINARY; PRT; 449 AA.
AC Q9KHF9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Outer membrane protein P1.
GN OMPP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-105;
RA MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL; AF260370; AAF97584.1; -
SQ SEQUENCE 449 AA; 48526 MW; 693B515B2112B536 CRC64;

```

Query Match 7.3%; Score 13; DB 2; Length 449;  
 Best Local Similarity 100.0%; Pred. No. 6.8e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133

Db 390 GATYKFTPNLSVD 402

```

RESULT 4
Q00754
ID Q00754 PRELIMINARY; PRT; 449 AA.
AC Q00754;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE Outer membrane protein P1 precursor (OMP P1).
GN OMP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90035394; PubMed=2572549;
RA Munson R.S. Jr., Grass S., Einhorn M., Bailey C., Newell C.;
RT "Comparative analysis of the structures of the outer membrane protein
RT P1 genes from major clones of Haemophilus influenzae type b.";
RL Infect. Immun. 57:3300-3305(1989).
DR EMBL: W73494; AAA24992.1;
KW Outer membrane; Signal; Transmembrane.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 449 OUTER MEMBRANE PROTEIN P1.
SQ SEQUENCE 449 AA; 48315 MW; 29C668F90AEC03F9 CRC64;

Query Match 7.3%; Score 13; DB 2; Length 449;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133
DB 390 GATYKFTPNLSVD 402
|||||

RESULT 5
Q00754
ID Q00754 PRELIMINARY; PRT; 450 AA.
AC Q00754;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein P1.
GN OMP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260375; AAF97589.1;
KW Haemophilus influenzae.
FT CHAIN 23 449 OUTER MEMBRANE PROTEIN P1.
SQ SEQUENCE 450 AA; 48611 MW; B616AAB6FEF0252A CRC64;

Query Match 7.3%; Score 13; DB 2; Length 450;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133
DB 392 GATYKFTPNLSVD 404
|||||

RESULT 6
Q00754
ID Q00754 PRELIMINARY; PRT; 451 AA.
AC Q00754;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein P1.
GN OMP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260375; AAF97589.1;
KW Haemophilus influenzae.
FT CHAIN 23 449 OUTER MEMBRANE PROTEIN P1.
SQ SEQUENCE 451 AA; 48611 MW; B616AAB6FEF0252A CRC64;

Query Match 7.3%; Score 13; DB 2; Length 451;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133
DB 391 GATYKFTPNLSVD 403
|||||

RESULT 7
Q00754
ID Q00754 PRELIMINARY; PRT; 451 AA.
AC Q00754;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein P1.
GN OMP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260375; AAF97589.1;
KW Haemophilus influenzae.
FT CHAIN 23 449 OUTER MEMBRANE PROTEIN P1.
SQ SEQUENCE 451 AA; 48611 MW; B616AAB6FEF0252A CRC64;

Query Match 7.3%; Score 13; DB 2; Length 451;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133
DB 392 GATYKFTPNLSVD 404
|||||

RESULT 8
Q00754
ID Q00754 PRELIMINARY; PRT; 451 AA.
AC Q00754;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein P1.
GN OMP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260375; AAF97589.1;
KW Haemophilus influenzae.
FT CHAIN 23 449 OUTER MEMBRANE PROTEIN P1.
SQ SEQUENCE 451 AA; 48611 MW; B616AAB6FEF0252A CRC64;

Query Match 7.3%; Score 13; DB 2; Length 451;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133
DB 391 GATYKFTPNLSVD 403
|||||
```

```

DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE Outer membrane protein P1.
GN OMPP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=658;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RT "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL; AF260367; AAF97581.1; -
SQ SEQUENCE 451 AA; 48640 MW; 72C6977C6F2A7BA1 CRC64;

Query Match 7.3%; Score 13; DB 2; Length 451;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 GATYKFTPNLSVD 133
Db 391 GATYKFTPNLSVD 403

RESULT 9
Q9KHF8 PRELIMINARY; PRT; 451 AA.
ID Q9KHF8
AC Q9KHF8;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE Outer membrane protein P1.
GN OMPP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2-H-1038;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RT "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL; AF260371; AAF97581.1; -
SQ SEQUENCE 451 AA; 48631 MW; E9D7DF9390C0BDB CRC64;

Query Match 7.3%; Score 13; DB 2; Length 451;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 GATYKFTPNLSVD 133
Db 392 GATYKFTPNLSVD 404

RESULT 10
Q9KHF7 PRELIMINARY; PRT; 451 AA.
ID Q9KHF7
AC Q9KHF7;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE Outer membrane protein P1.
GN OMPP1.
OS Haemophilus influenzae.

```

```

DE Outer membrane protein P1.
GN OMPP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=658;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RT "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL; AF260372; AAF97586.1; -
SQ SEQUENCE 451 AA; 48796 MW; 3C6631D057EB0FFB CRC64;

Query Match 7.3%; Score 13; DB 2; Length 451;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 GATYKFTPNLSVD 133
Db 392 GATYKFTPNLSVD 404

RESULT 11
Q9KHF6 PRELIMINARY; PRT; 451 AA.
ID Q9KHF6
AC Q9KHF6;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE Outer membrane protein P1.
GN OMPP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13-H-1157;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RT "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL; AF260373; AAF97587.1; -
SQ SEQUENCE 451 AA; 48446 MW; F4DD7C9354DFCF13 CRC64;

Query Match 7.3%; Score 13; DB 2; Length 451;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 GATYKFTPNLSVD 133
Db 392 GATYKFTPNLSVD 404

RESULT 12
Q9KHG1 PRELIMINARY; PRT; 454 AA.
ID Q9KHG1
AC Q9KHG1;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Outer membrane protein P1.
GN OMPP1.
OS Haemophilus influenzae.

```



```
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22-H-1154;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RT Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260366; AAF97562.1; -.
SQ SEQUENCE 454 AA; 48986 MW; 494FBAC2600DA2C8 CRC64;

Query Match 7.3%; Score 13; DB 2; Length 454;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133
|
Db 392 GATYKFTPNLSVD 404

RESULT 13
Q9KHH5 PRELIMINARY; PRT; 455 AA.
AC Q9KHH5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=7004;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RT Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260348; AAF97562.1; -.
SQ SEQUENCE 455 AA; 49165 MW; 280E211B16C18C0 CRC64;

Query Match 7.3%; Score 13; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133
|
Db 393 GATYKFTPNLSVD 405

RESULT 14
Q9KHH4 PRELIMINARY; PRT; 455 AA.
AC Q9KHH4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6094;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RT Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260349; AAF97563.1; -.
SQ SEQUENCE 455 AA; 49111 MW; 0CB00B162A6A6CB1 CRC64;

Query Match 7.3%; Score 13; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133
|
Db 393 GATYKFTPNLSVD 405

RESULT 15
Q9KHH3 PRELIMINARY; PRT; 455 AA.
AC Q9KHH3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC9795;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RT Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260352; AAF97566.1; -.
SQ SEQUENCE 455 AA; 49169 MW; 77A8BA97BCE2B450 CRC64;

Query Match 7.3%; Score 13; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133
|
Db 393 GATYKFTPNLSVD 405

RESULT 16
Q9KHH2 PRELIMINARY; PRT; 455 AA.
AC Q9KHH2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=161;
```

```

RX MEDLINE-20359342; PubMed-10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL; AF260353; AAF97567.1; -.
SQ SEQUENCE 455 AA; 49139 MW; 9CCB2EE675B421D9 CRC64;

Query Match          7.3%; Score 13; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133
DB 393 GATYKFTPNLSVD 405

RESULT 17
Q9KHH1 ID Q9KHH1 PRELIMINARY; PRT; 455 AA.
AC Q9KHH1;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE Outer membrane protein p1.
GN OMPP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-OT-9;
RX MEDLINE-20359342; PubMed-10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL; AF260354; AAF97568.1; -.
SQ SEQUENCE 455 AA; 49123 MW; FD5882979F10EB24 CRC64;

Query Match          7.3%; Score 13; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133
DB 393 GATYKFTPNLSVD 405

RESULT 18
Q00450 ID Q00450 PRELIMINARY; PRT; 455 AA.
AC Q00450;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE Outer membrane protein p1 precursor, subtype 3L (OMP p1).
GN OMPP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-1613 / SEROTYPE B;
RX MEDLINE-90035394; PubMed-2572549;
RA Munson R.S. Jr., Grass S., Einhorn M., Bailey C., Newell C.;
RA "Comparative analysis of the structures of the outer membrane protein

```

```

RT Pl genes from major clones of Haemophilus influenzae type b.";
RL Infect. Immun. 57:3300-3305(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-7109, AND 6107;
RX MEDLINE-20359342; PubMed-10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: STRONG, TO E.COLI LONG-CHAIN FATTY ACID TRANSPORT
CC PROTEIN (PADL).
DR EMBL; M63151; AAA24991.1; -.
DR EMBL; AF260351; AAF97565.1; -.
DR EMBL; AF260350; AAF97564.1; -.
DR PIR; A40183; A40183.
KW Outer membrane; Signal; Transmembrane.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 455 OUTER MEMBRANE PROTEIN p1.
SQ SEQUENCE 455 AA; 49141 MW; 795059130A39D9E1 CRC64;

Query Match          7.3%; Score 13; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133
DB 393 GATYKFTPNLSVD 405

RESULT 19
Q9KH14 ID Q9KH14 PRELIMINARY; PRT; 456 AA.
AC Q9KH14;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE Outer membrane protein p1.
GN OMPP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BCH-1;
RX MEDLINE-20359342; PubMed-10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL; AF260337; AAF97551.1; -.
SQ SEQUENCE 456 AA; 49402 MW; 8CF3E87CB810D6E6 CRC64;

Query Match          7.3%; Score 13; DB 2; Length 456;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133
DB 394 GATYKFTPNLSVD 406

RESULT 20
Q9KH13 ID Q9KH13 PRELIMINARY; PRT; 456 AA.
AC Q9KH13;
DT 01-OCT-2000 (TRENBLrel. 15, Created)

```



```

OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=27-H-1433;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RT Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260342; AAF97556.1; -.
SQ SEQUENCE 456 AA; 49239 MW; 35101EFC7A6769D2 CRC64;

Query Match 7.3%; Score 13; DB 2; Length 456;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 GATYKFTPNLSVD 133
Db 394 GATYKFTPNLSVD 406

RESULT 25
Q9KHG6 ID Q9KHG6 PRELIMINARY; PRT; 456 AA.
AC Q9KHG6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein P1.
GN OMPP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=7416;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RT Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260347; AAF97561.1; -.
SQ SEQUENCE 456 AA; 49370 MW; 0FA50D60942905D0 CRC64;

Query Match 7.3%; Score 13; DB 2; Length 456;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 GATYKFTPNLSVD 133
Db 394 GATYKFTPNLSVD 406

RESULT 26
Q9KHG9 ID Q9KHG9 PRELIMINARY; PRT; 456 AA.
AC Q9KHG9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein P1.
GN OMPP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LO-127;
RX MEDLINE=20359342; PubMed=10899849;

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=1071;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RT Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260358; AAF97572.1; -.
SQ SEQUENCE 456 AA; 49130 MW; C16D7F3809AC8FA2 CRC64;

Query Match 7.3%; Score 13; DB 2; Length 456;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 GATYKFTPNLSVD 133
Db 394 GATYKFTPNLSVD 406

RESULT 27
Q9KHG8 ID Q9KHG8 PRELIMINARY; PRT; 456 AA.
AC Q9KHG8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein P1.
GN OMPP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BCH-2;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RT Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein P1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260359; AAF97573.1; -.
SQ SEQUENCE 456 AA; 49075 MW; CF7B3ECB4213C841 CRC64;

Query Match 7.3%; Score 13; DB 2; Length 456;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 GATYKFTPNLSVD 133
Db 394 GATYKFTPNLSVD 406

RESULT 28
Q9KHG7 ID Q9KHG7 PRELIMINARY; PRT; 456 AA.
AC Q9KHG7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Outer membrane protein P1.
GN OMPP1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LO-127;
RX MEDLINE=20359342; PubMed=10899849;

```

```

RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RT Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260360; AAF97574.1; -.
SQ SEQUENCE 456 AA; 49220 MW; 7AC1F56499D94C91 CRC64;

Query Match 7.3%; Score 13; DB 2; Length 456;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133
Db 394 GATYKFTPNLSVD 406

RESULT 29
Q9KHG5 PRELIMINARY; PRT; 456 AA.
AC Q9KHG5;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1512A;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260362; AAF97576.1; -.
SQ SEQUENCE 456 AA; 49044 MW; 3581AE619228511E CRC64;

Query Match 7.3%; Score 13; DB 2; Length 456;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133
Db 395 GATYKFTPNLSVD 407

RESULT 30
Q9KHG4 PRELIMINARY; PRT; 456 AA.
AC Q9KHG4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=49;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a

```

```

RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260363; AAF97577.1; -.
SQ SEQUENCE 456 AA; 49253 MW; A39BBE8891302566 CRC64;

Query Match 7.3%; Score 13; DB 2; Length 456;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133
Db 395 GATYKFTPNLSVD 407

RESULT 31
Q9KHG3 PRELIMINARY; PRT; 456 AA.
AC Q9KHG3;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1-H-1085;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).
DR EMBL: AF260364; AAF97578.1; -.
SQ SEQUENCE 456 AA; 49133 MW; OD6AFC5DFEB35014 CRC64;

Query Match 7.3%; Score 13; DB 2; Length 456;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133
Db 395 GATYKFTPNLSVD 407

RESULT 32
Q9KHG6 PRELIMINARY; PRT; 457 AA.
AC Q9KHG6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE Outer membrane protein p1.
GN OMPp1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5;
RX MEDLINE=20359342; PubMed=10899849;
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,
RA Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;
RT "Variability of outer membrane protein p1 and its evaluation as a
RT vaccine candidate against experimental otitis media due to nontypeable
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";
RL Infect. Immun. 68:4505-4517(2000).

```

DR EMBL; AF260361; AAF97575.1; -.  
SQ SEQUENCE 457 AA; 49292 MW; D25CB4D6E81A965F CRC64;

Query Match 7.3%; Score 13; DB 2; Length 457;  
Best Local Similarity 100.0%; Pred. No. 7e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133  
| | | | | | | | | |  
DB 395 GATYKFTPNLSVD 407

RESULT 33  
Q9K397 ID Q9K397 PRELIMINARY; PRT; 458 AA.  
AC Q9K397;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE Outer membrane protein PI.  
GN OMPPI.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OT-81, AND 1363;  
RX MEDLINE=20359342; PubMed=10899849;  
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,  
RT "Variability of outer membrane protein PI and its evaluation as a  
RT vaccine candidate against experimental otitis media due to nontypeable  
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";  
RL Infect. Immun. 68:4505-4517(2000).  
DR EMBL; AF260356; AAF97570.1; -.  
DR EMBL; AF260355; AAF97569.1; -.  
SQ SEQUENCE 458 AA; 49674 MW; 1F886C1FAD682F94 CRC64;

Query Match 7.3%; Score 13; DB 2; Length 458;  
Best Local Similarity 100.0%; Pred. No. 7e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133  
| | | | | | | | | |  
DB 396 GATYKFTPNLSVD 408

RESULT 34  
Q9K324 ID Q9K324 PRELIMINARY; PRT; 459 AA.  
AC Q9K324;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE Outer membrane protein PI.  
GN OMPPI.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=218, AND 200;  
RX MEDLINE=20359342; PubMed=10899849;  
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,  
RT "Variability of outer membrane protein PI and its evaluation as a  
RT vaccine candidate against experimental otitis media due to nontypeable  
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";  
RL Infect. Immun. 68:4505-4517(2000).  
DR EMBL; AF260344; AAF97558.1; -.  
DR EMBL; AF260343; AAF97557.1; -.  
SQ SEQUENCE 459 AA; 49674 MW; 1F886C1FAD682F94 CRC64;

SQ SEQUENCE 459 AA; 49717 MW; 54490C08E4729870 CRC64;

Query Match 7.3%; Score 13; DB 2; Length 459;  
Best Local Similarity 100.0%; Pred. No. 7e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133  
| | | | | | | | | |  
DB 397 GATYKFTPNLSVD 409

RESULT 35  
Q9KH8 ID Q9KH8 PRELIMINARY; PRT; 459 AA.  
AC Q9KH8;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE Outer membrane protein PI.  
GN OMPPI.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=667;  
RX MEDLINE=20359342; PubMed=10899849;  
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,  
RT "Variability of outer membrane protein PI and its evaluation as a  
RT vaccine candidate against experimental otitis media due to nontypeable  
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";  
RL Infect. Immun. 68:4505-4517(2000).  
DR EMBL; AF260345; AAF97559.1; -.  
SQ SEQUENCE 459 AA; 49773 MW; 5E6DCD78F9A8FF8A CRC64;

Query Match 7.3%; Score 13; DB 2; Length 459;  
Best Local Similarity 100.0%; Pred. No. 7e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133  
| | | | | | | | | |  
DB 397 GATYKFTPNLSVD 409

RESULT 36  
Q9KH7 ID Q9KH7 PRELIMINARY; PRT; 459 AA.  
AC Q9KH7;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Outer membrane protein PI.  
GN OMPPI.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=199;  
RX MEDLINE=20359342; PubMed=10899849;  
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,  
RT "Variability of outer membrane protein PI and its evaluation as a  
RT vaccine candidate against experimental otitis media due to nontypeable  
RT Haemophilus influenzae: an unambiguous, multifaceted approach.";  
RL Infect. Immun. 68:4505-4517(2000).  
DR EMBL; AF260346; AAF97560.1; -.  
SQ SEQUENCE 459 AA; 49745 MW; 5A46FF07187E9883 CRC64;

Query Match 7.3%; Score 13; DB 2; Length 459;

Best Local Similarity 100.0%; Pred. No. 7e-05; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133  
|||||

Db 397 GATYKFTPNLSVD 409

## RESULT 37

Q9KHH0 PRELIMINARY; PRT; 460 AA.  
AC Q9KHH0; 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DE Outer membrane protein PI.  
GN OMP1.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
RN NCBI\_TaxID=727;  
RP SEQUENCE FROM N.A.  
RC STRAIN=9-H-1194;  
RX MEDLINE=20359342; PubMed=10899849;  
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,  
Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;  
RT "Variability of outer membrane protein PI and its evaluation as a  
vaccine candidate against experimental otitis media due to nontypeable  
Haemophilus influenzae: an unambiguous, multifaceted approach."  
RL Infect. Immun. 68:4503-4517(2000).  
DR EMBL: AF260357; AAF97588.1;  
SQ SEQUENCE 460 AA; 49719 MW; F1CEBEB597A5CBA7 CRC64;

Query Match 7.3%; Score 13; DB 2; Length 460;  
Best Local Similarity 100.0%; Pred. No. 7e-05; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133  
|||||

Db 398 GATYKFTPNLSVD 410

## RESULT 38

Q9KHF5 PRELIMINARY; PRT; 449 AA.  
AC Q9KHF5; 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DE Outer membrane protein PI.  
GN OMP1.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
RN NCBI\_TaxID=727;  
RP SEQUENCE FROM N.A.  
RC STRAIN=9;  
RX MEDLINE=20359342; PubMed=10899849;  
RA Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,  
Truong-Bolduc Q.C., Rice P.A., Pelton S.I., Goldstein R.;  
RT "Variability of outer membrane protein PI and its evaluation as a  
vaccine candidate against experimental otitis media due to nontypeable  
Haemophilus influenzae: an unambiguous, multifaceted approach."  
RL Infect. Immun. 68:4503-4517(2000).  
DR EMBL: AF260374; AAF97588.1;  
SQ SEQUENCE 449 AA; 48424 MW; F536A974B0C7F827 CRC64;

Query Match 5.6%; Score 10; DB 2; Length 449;  
Best Local Similarity 100.0%; Pred. No. 0.095; Mismatches 0; Indels 0; Gaps 0;

QY 121 GATYKFTPNLSVD 133  
|||||

Db 398 GATYKFTPNLSVD 410

## RESULT 39

Q9KLF5 PRELIMINARY; PRT; 412 AA.  
AC Q9KLF5; 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DE Outer membrane protein PI.  
GN OMP1.  
OS Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
RN NCBI\_TaxID=666;  
RP SEQUENCE FROM N.A.  
RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
RX MEDLINE=20406833; PubMed=10952301;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
cholerae."  
RL Nature 406:477-483(2000).  
DR EMBL: AE004414; AAF96760.1;  
DR TIGR: VCA0862;  
DR InterPro: IPR001064; Crystallin.  
DR PROSITE: PS00225; CRYSTALLIN\_BETAGAMMA; UNKNOWN\_1.  
KW Complete proteome.  
SQ SEQUENCE 412 AA; 44784 MW; FB606C929838D891 CRC64;

Query Match 5.0%; Score 9; DB 16; Length 412;  
Best Local Similarity 100.0%; Pred. No. 0.98; Mismatches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 WELSGFHL 31  
|||||

Db 261 WELSGFHL 269

## RESULT 40

O28679 PRELIMINARY; PRT; 72 AA.  
AC O28679; 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE Heme D1 biosynthesis protein (NIRD).  
GN AF1593.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
RN NCBI\_TaxID=2234;  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,  
Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,  
Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,

Query Match 5.0%; Score 9; DB 16; Length 412;  
Best Local Similarity 100.0%; Pred. No. 0.98; Mismatches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 WELSGFHL 31  
|||||

Db 261 WELSGFHL 269

## RESULT 41

O28679 PRELIMINARY; PRT; 72 AA.  
AC O28679; 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE Heme D1 biosynthesis protein (NIRD).  
GN AF1593.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
RN NCBI\_TaxID=2234;  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,  
Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,  
Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,

Query Match 5.0%; Score 9; DB 16; Length 412;  
Best Local Similarity 100.0%; Pred. No. 0.98; Mismatches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 WELSGFHL 31  
|||||

Db 261 WELSGFHL 269

```

RA  Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA  Venter J.C.;
RT  "The complete genome sequence of the hyperthermophilic, sulphate-
RL  reducing archaeon Archaeoglobus fulgidus."
RL  Nature 390:364-370(1997).
DR  EMBL: AE000992; AAB89654.1; -.
DR  TIGR: AF1593; -.
DR  InterPro: IPR003662; sub_transporter.
DR  PROSITE: PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 72 AA; 8211 MW; 6FBA1C8802970079 CRC64;

Query Match      4.5%  Score 8;  DB 17;  Length 72;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  171 YGLNLNYR 178
    |||||
Db  52 YGLNLNYR 59

Search completed: May 12, 2003, 09:48:49
Job time : 70 secs

```



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 19, 2003, 18:57:34 ; Search time 1734'Seconds  
(without alignments)  
3004.269 Million cell updates/sec

Title: US-09-995-493-52

Perfect score: 957

Sequence: 1 QHNGVLGPYIGKSLTKLP.....ADYTKATANLYGLNLNRYR 179

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-O=/cnp2.1/USPTO.spool/US0995493/runat\_12052003\_091308\_23003/app\_query.fasta\_1.327  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -NORM=ext -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0995493.ecgn.1.1.1687@runat\_12052003\_091308\_23003 -NCPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg\_hum.\*  
31: em.htg\_inv.\*  
32: em.htg\_other.\*  
33: em.htg\_mus.\*  
34: em.htg\_pln.\*  
35: em.htg\_rtd.\*  
36: em.htg\_mam.\*  
37: em.htg\_vrt.\*  
38: em.sy.\*  
39: em.htgo\_hum.\*  
40: em.htgo\_mus.\*  
41: em.htgo\_other.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	580.5	60.7	1284	1 HIU28808	U28808 Haemophilus
2	571.5	59.7	1350	1 AF260368	AF260368 Haemophilus
3	571.5	59.7	1350	1 AF260369	AF260369 Haemophilus
4	571.5	59.7	1350	1 AF260370	AF260370 Haemophilus
5	570.5	59.6	1347	6 AR085993	AR085993 Sequence
6	570.5	59.6	1350	1 HEAOMPP1B	M73494 Haemophilus
7	564	58.9	1356	1 AF260371	AF260371 Haemophilus
8	563.5	58.9	1356	1 AF260372	AF260372 Haemophilus
9	562.5	58.8	1350	1 AF260374	AF260374 Haemophilus
10	561	58.6	1368	1 AF260353	AF260353 Haemophilus
11	561	58.6	1368	1 AF260354	AF260354 Haemophilus
12	559.5	58.5	1371	1 AF260364	AF260364 Haemophilus
13	559	58.4	1365	6 AR085992	AR085992 Sequence
14	559	58.4	1368	1 AF260349	AF260349 Haemophilus
15	559	58.4	1368	1 AF260350	AF260350 Haemophilus
16	559	58.4	1368	1 AF260351	AF260351 Haemophilus
17	559	58.4	1368	1 AF260352	AF260352 Haemophilus
18	559	58.4	1368	1 HEAOMPE1A	M63151 Haemophilus
19	557	58.2	1368	1 AF260348	AF260348 Haemophilus
20	553.5	57.8	1356	1 AF260373	AF260373 Haemophilus
21	552.5	57.7	1353	1 AF260375	AF260375 Haemophilus
22	552.5	57.7	1353	1 AF260376	AF260376 Haemophilus
23	550.5	57.5	1356	1 AF260367	AF260367 Haemophilus
24	549.5	57.4	1371	1 AF260358	AF260358 Haemophilus
25	549	57.4	1371	1 AF260347	AF260347 Haemophilus
26	549	57.4	1371	1 AF260360	AF260360 Haemophilus
27	549	57.4	1381	6 AR199524	AR199524 Sequence
28	549	57.4	1560	6 AR085991	AR085991 Sequence
29	549	57.4	1598	1 HEAOMPP1	J03381 Haemophilus
30	549	57.4	1598	6 AR089267	AR089267 Sequence
31	549	57.4	1598	6 AR093467	AR093467 Sequence
32	549	57.4	13070	1 U32723	U32723 Haemophilus
33	547	57.2	1356	1 AF260365	AF260365 Haemophilus
34	547	57.2	1380	1 AF260343	AF260343 Haemophilus
35	547	57.2	1380	1 AF260344	AF260344 Haemophilus
36	547	57.2	1380	1 AF260345	AF260345 Haemophilus
37	547	57.2	1380	1 AF260346	AF260346 Haemophilus
38	545.5	57.0	1371	1 AF260359	AF260359 Haemophilus
39	544	56.8	1371	1 AF260337	AF260337 Haemophilus
40	544	56.8	1371	1 AF260339	AF260339 Haemophilus
41	542.5	56.7	1374	1 AF260361	AF260361 Haemophilus
42	539.5	56.4	14766	1 AE006147	AE006147 Pasteurel
43	538	56.2	1371	1 AF260342	AF260342 Haemophilus
44	538	56.2	1377	1 AF260355	AF260355 Haemophilus
45	538	56.2	1377	1 AF260356	AF260356 Haemophilus

ALIGNMENTS

<hr/>					
Percent Similarity:	82.14%	Conservative:	31		
Best Local Similarity:	63.69%	Mismatches:	27		
Query Match:	60.66%	Indels:	3		
DB:	1	Gaps:	2		
<hr/>					
US-09-995-493-52 (1-179) x HIU28808 (1-1284)					
Qy	12	LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluleuSerGlyPheHisGlnLeu	31		
Dd	787	AAAGGTATTAAACCTTTAATTCGCCAGATTACTAGAACCCTTCAGGTTTTCCATCAATTA	846		
Qy	32	ThrAspGlnTrpAlaIleHisTySerTyTLystYrThrGluTrpSerArgPheLysGlu	51		
Dd	847	ACTCACAAAATTGCTGGCATATAGTTATAATATACCCATTGGAGTCGTTTAAACAAAA	906		
Qy	52	LeuArgGlyLysTyGlnAspGlySerGlyTyTGluAlaPheThrlLysLysGluGluTyr	71		
Dd	907	TTCATCGCTAGCTCGAAGATGGTAA-- --AAAGCGTTTGATAAAGAATTACAATAC	960		
Qy	72	LysAspAsnSerArgPheAlaIleGlyThrTyTSerLeuAsnAspAlaLeuThrLeu	91		
Dd	961	AGTAATAACTCTCGTTGGTAGGGGAAGTTATAATCTTGATGAAAAATTGACCTTA	1020		
Qy	92	ArgAlaGlyLeuAlaTyAspLysAlaLaSerLysThrHisLeuSerAlaSerIlePro	111		
Dd	1021	CGTGCGGGTATTGCTGGGATCAAGCGGCATCTCGTCATCACCCTAGTGTGCAATTCCA	1080		
Qy	112	AspThrAspArgMetTrpTySerIleGlyAlaThrTyrlLysPheThrProAsnLeuSer	131		
Dd	1081	GATACCGATCGCACTTGGTATATCTTAGGGGCAACCTATAAATTCACGCGGAATTTATCT	1140		
Qy	132	ValAspValGlyPheAlaHisLeuArgGlyLysLysHisPheValGluThrGlnAsn	151		
Dd	1141	GTTGATCTTGGCTATGCTTACTTAAAGCGAAAAAGTTCACITTTAAGAAAGTACAACAA	1200		
Qy	152	IleLySGlyLeuLeuLeuValGluAlaAspTyThrThrLysAlaThrAlaAsnLeuTyr	171		
Dd	1201	GTT---GGCTCATATAACAACCGCAAATTTACACTTCTCAAGCACACGCAATCTTTAC	1257		
Qy	172	GlyLeuAsnLeuAsnTyArgPhe	179		
Dd	1258	GGCTTAACTTAAATTATATGTTTC	1281		
<hr/>					
RESULT 2					
AF260368	1350 bp DNA linear BCT 07-AUG-2000				
LOCUS	Haemophilus influenzae strain 6255 outer membrane protein P1				
DEFINITION	(ompP1) gene, complete cds.				
ACCESSION	AF260368				
VERSION	AF260368.1 GI:9716628				
KEYWORDS	Haemophilus influenzae.				
SOURCE	Haemophilus influenzae				
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;				
	Haemophilus.				
REFERENCE	1 (bases 1 to 1350)				
AUTHORS	Bolduc,G.R., Bouchet,V., Jiang,R.Z., Geisselsoder,J., Truong-Bolduc,Q.C., Rice,P.A., Pelton,S.I. and Goldstein,R. Variability of outer membrane protein P1 and its evaluation as a vaccine candidate against experimental otitis media due to nontypeable Haemophilus influenzae: an unambiguous, multifaceted approach				
JOURNAL	Infect. Immun. 68 (8), 4505-4517 (2000)				
MEDLINE	20359342				
PUBMED	10899849				
REFERENCE	2 (bases 1 to 1350)				
AUTHORS	Bouchet,V.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-APR-2000) Section of Molecular Genetics, The Maxwell Finland Laboratory for Infectious Diseases, Boston University School of Medicine, Boston Medical Center, 774 Albany Street, Boston, MA 02118, USA				
FEATURES	Location/Qualifiers				

```
source
1. .1350
/organism="Haemophilus influenzae"
/strain="6255"
/serotype="non-typable"
/db_xref="taxon:727"
/country="United Kingdom"
1. .1350
/genes="omp1"
1. .1350
/genes="omp1"
/notes="OMP p1"
/codon_start=1
/transl_table=11
/product="outer membrane protein p1"
/protein_id="AAF97582.1"
/db_xref="GI:9716629"
/transl_table="MKKFNQSLATAMLLAAGGANAFAQLAEVSTSGIGRAYAGEAA
IADNASVATNPALMSLEKTAQSTGGVYVDSIRNNMGVDSSITATMRTTKYGSAS
ARNVPGAEVPLNYFVAPVNDKRALGAGNMFGLSKSYDSDSITAGVFGGKDLTAIN
LNLGSAVRTEGLSLGLGVNAVYAKAQRNAGIIAESVKIAQNAIKTVNPKDKATDY
LTSKDSVSLQDRAAGWGFNAGVYQFNEANRIGLAYHSKVDIDFTDRTATSEAE
VTEAGKGNLTLLTPDYLELSGFHQLTKFAVHYSYKTHWSRLTKLHASFDGKKAF
DKELQYNNRSVALGASYNLDEKLTIRAGIAYDOAASRHRSAAIPTDRTWYSLGAT
YKFTPNLSVDLGLVAYLKGGKHVFEKQKAVGGFITTTANYTSQAHNLYGLNLNYSF"
BASE COUNT 413 a 233 c 294 g 410 t
ORIGIN

Alignment Scores:
Pred. No.: 1.16e-50 Length: 1350
Score: 571.50 Matches: 106
Percent Similarity: 80.47% Conservativeness: 30
Best Local Similarity: 62.72% Mismatches: 30
Query Match: 59.72% Indels: 3
DB: 1 Gaps: 2

US-09-995-493-52 (1-179) x AF260368 (1-1350)
QY 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31
Db 847 AAAGGTAATTAACCTTACATGCCAGATTACTTAGAACCTTCTGTTCCATCAATTA 906
QY 32 ThrAspGlnTrpAlaLeuHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51
Db 907 ACTGACAAATTTTCAGTGCATATAGTTATAAATATACCCATTGGAGTCTTTTAAACAAA 966
QY 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71
Db 967 TTACATGCTAGCTTCGAGATGCTAG-----AAAGCTTTTCATTAAGAAATTCGAATAC 1020
QY 72 LysAspAsnSerArgPheAlaLeuGlyThrTyrSerLeuAsnAspAlaLeuThrLeu 91
Db 1021 AGTAATAACTCTCGTGTGCTAGGCGCAAGTTATATCTTATGATGAAAAATTAACCTTA 1080
QY 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111
Db 1081 CGTGGCGGATGCTGTCACGATCAAGTGCATCTCGCATCACCGTACGCTGCAATTCCTCA 1140
QY 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131
Db 1141 GATACCGATCGCACTGGTATAGTTAGGGCGCAACCTATAAATTCACCGCAATTTATCT 1200
QY 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151
Db 1201 GTTGATCTTGGCTATGCTTACTTAAAGGCAAAAGATTCACCTTAAAGAAAGTACAAAA 1260
QY 152 ---IleLysGlyLeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeu 170
Db 1261 GCTGAGGTGGCTTCAACAAACCGCAAAATACACTTCTCAAGCACACCGCAAAATCTT 1320
QY 171 TyrGlyLeuAsnLeuAsnTyrArgPhe 179
Db 1321 TACGGCTTAAACTTAAATATATAGTTTC 1347
```

```
RESULT 3
AF260369
LOCUS
DEFINITION
Haemophilus influenzae strain ATCC9833 outer membrane protein p1
(ompP1) gene, complete cds.
ACCESSION
AF260369
VERSION
AF260369.1 GI:9716630
SOURCE
Haemophilus influenzae.
ORGANISM
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
REFERENCE
1 (bases 1 to 1350)
AUTHORS
Bolduc.G.R., Bouchet.V., Jiang.R.Z., Geiselsoder.J.,
Truong-Bolduc.Q.C., Rice.P.A., Pelton.S.I. and Goldstein.R.
TITLE
Variability of outer membrane protein p1 and its evaluation as a
vaccine candidate against experimental otitis media due to
nontypeable Haemophilus influenzae: an unambiguous, multifaceted
approach
JOURNAL
Infect. Immun. 68 (8), 4505-4517 (2000)
MEDLINE
20359342
PUBMED
10899849
REFERENCE
2 (bases 1 to 1350)
AUTHORS
Bouchet.V.
TITLE
Direct Submission
JOURNAL
Submitted (25-APR-2000) Section of Molecular Genetics, The Maxwell
Finland Laboratory for Infectious Diseases, Boston University
School of Medicine, Boston Medical Center, 774 Albany Street,
Boston, MA 02118, USA
FEATURES
Location/Qualifiers
1. .1350
/organism="Haemophilus influenzae"
/strain="ATCC9833"
/serotype="f"
/db_xref="taxon:727"
1. .1350
/genes="omp1"
1. .1350
/genes="omp1"
/notes="OMP p1"
/codon_start=1
/transl_table=11
/product="outer membrane protein p1"
/protein_id="AAF97583.1"
/db_xref="GI:9716631"
/transl_table="MKKFNQSLATAMLLAAGGANAFAQLAEVSTSGIGRAYAGEAA
IADNASVATNPALMSLEKTAQSTGGVYVDSIRNNMGVDSSITATMRTTKYGSAS
ARNVPGAEVPLNYFVAPVNDKRALGAGNMFGLSKSYDSDSITAGVFGGKDLTAIN
LNLGSAVRTEGLSLGLGVNAVYAKAQRNAGIIAESVKIAQNAIKTVNPKDKATDY
LTSKDSVSLQDRAAGWGFNAGVYQFNEANRIGLAYHSKVDIDFTDRTATSEAE
VTEAGKGNLTLLTPDYLELSGFHQLTKFAVHYSYKTHWSRLTKLHASFDGKKAF
DKELQYNNRSVALGASYNLDEKLTIRAGIAYDOAASRHRSAAIPTDRTWYSLGAT
YKFTPNLSVDLGLVAYLKGGKHVFEKQKAVGGFITTTANYTSQAHNLYGLNLNYSF"
BASE COUNT 414 a 232 c 294 g 410 t
ORIGIN

Alignment Scores:
Pred. No.: 1.16e-50 Length: 1350
Score: 571.50 Matches: 106
Percent Similarity: 80.47% Conservativeness: 30
Best Local Similarity: 62.72% Mismatches: 30
Query Match: 59.72% Indels: 3
DB: 1 Gaps: 2

US-09-995-493-52 (1-179) x AF260369 (1-1350)
QY 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31
Db 847 AAAGGTAATTAACCTTACATGCCAGATTACTTAGAACCTTCTGTTCCATCAATTA 906
QY 32 ThrAspGlnTrpAlaLeuHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51
Db 907 ACTGACAAATTTTCAGTGCATATAGTTATAAATATACCCATTGGAGTCTTTTAAACAAA 966
```





```

|||||
Db 1321 TACGGCTAAACTTAAATATTAGTTTC 1347
RESULT 7
AF260371
LOCUS AF260371 1356 bp DNA linear BCT 07-AUG-2000
DEFINITION Haemophilus influenzae strain 2-H-1038 outer membrane protein p1
ACCESSION AF260371
VERSION AF260371.1 GI:9716634
KEYWORDS Haemophilus influenzae.
SOURCE Haemophilus influenzae.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
REFERENCE 1 (bases 1 to 1356)
AUTHORS Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder, J.,
Truong-Bolduc, O.C., Rice, P.A., Pelton, S.I., and Goldstein, R.
TITLE Variability of outer membrane protein p1 and its evaluation as a
vaccine candidate against experimental otitis media due to
nontypeable Haemophilus influenzae: an unambiguous, multifaceted
approach
JOURNAL Infect. Immun. 68 (8), 4505-4517 (2000)
MEDLINE 20359342
PUBMED 10899849
REFERENCE 2 (bases 1 to 1356)
AUTHORS Bouchet, V.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2000) Section of Molecular Genetics, The Maxwell
Finland Laboratory for Infectious Diseases, Boston University
School of Medicine, Boston Medical Center, 774 Albany Street,
Boston, MA 02118, USA
FEATURES
source
1. .1356
/organism="Haemophilus influenzae"
/strain="2-H-1038"
/serotype="non-typable"
/db_xref="taxon:727"
/country="USA"
1. .1356
/gene="ompP1"
1. .1356
/gene="ompP1"
/feature="OMP P1"
/codon_start=1
/transl_table=11
/product="outer membrane protein p1"
/protein_id="AAF97585.1"
/db_xref="GI:9716635"
/translation="MKFQNSILATAMLLAAGGANAATQLAEVSTSGLRAYAGGAA
IADNASVATNPALMSLFQSTGGVYDSRINMGDVAASITGSSIKVTNDDVAS
ERNVPGFVFNLYFVAVNDKLVAGAGMNVNFKLSEYDSDYAGVFGKTDLTAIN
LNLGAYRVTEGLSLGLGVNAVYAKAQRVNAIGIADSVKPOAKRALETVPVPTKIP
DLYTSKDYVSLQDRAAGFCGWNAGVYQFNEANRIGLAYHSKVIDIDTDRPTSLG
KYGIEAGKTNLTFTLPDYLELSGFHQLTDKFAVHYSYKYTHWSRLTKLHSAVEDCKK
APDKELQYNNRSVALCASTNLEKUTLRAGIYDQNASRHRSAIPDPTDRTWISLG
F"
ATYKPTPNLSVDLGYAYLKGKVFHFVKVADDFITANANTYSQAHNLYGLNLNYS
F"
BASE COUNT 416 a 227 c 300 g 412 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 7.16e-50 Length: 1356
Score: 564.00 Matches: 108
Percent Similarity: 76.97% Conservative: 29
Best Local Similarity: 60.67% Mismatches: 37
Query Match: 58.93% Indels: 4
DB: 1 Gaps: 3
US-09-995-493-52 (1-179) x AF260371 (1-1356)
Qy 4 GlyValLeuGlyProTyrIleGlyLys---GlySerLeuThrLeuLysLeuProAlaTyr 22
|||||
|||||

```

```

/gene="ompP1"
/note="OMP P1"
/codon_start=1
/transl_table=11
/product="outer membrane protein p1"
/protein_id="AF97586.1"
/db_xref="GI:9716637"
/translation="MKKFNOSLLATAMLLAAGNAAAFOLAEVSTSGLRAYAGEAA
IADNSVATFNPALMSLFKTAQFSTGCVYDPRINMGDVNASKATMNTKYGSAQ
ERNVPCAFVFNLYFVAPNDKFAAGMNVNFKLSEYDDDDYDAGVGGKTDLSAIN
NLNSGAVRTTEGLSLGLGVNAVYAKAQVERNAGIIVRSVKDSOAOATLKVPTLIP
DLTSDKSVLSQDRAAWGFNAGVYQFNEANRIGLAYHSHKVIDDFDRTATSLG
OKSIEAGKTNLTLPDYLELSPGHQTLDFAVHYSYKTHWSRLTLPDYASTEDGKK
VFDELOYSNNRSVALGASYNLDEKLTLAGIAYDQAASRHRSAAIPDTRTWYSLG
ATYKFTPNLSVDLGAYLKGKKVHEKVKVADGLITTNANTSOAHANLYGLNLNYS
F"
BASE COUNT 415 a 237 c 295 g 409 t
ORIGIN
Alignment Scores:
Pred. No.: 8,08e-50 Length: 1356
Score: 563.50 Matches: 105
Percent Similarity: 79.64% Conservative: 28
Best Local Similarity: 62.87% Mismatches: 31
Query Match: 58.88% Indels: 3
DB: 1 Gaps: 2
US-09-995-493-52 (1-179) x AF260372 (1-1356)
QY 14 SerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAsp 33
Db 859 AATTAAACCCCTAGATTGCCAGATTACTTAGAACCTTCTGGTTCCATCAATTAACGAC 918
QY 34 GlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGluLeuArg 53
Db 919 AAATTCCTGCTGCATATAGTTATTAATATATACCATTCGGAGTCGTTTAACAGATTATAC 978
QY 54 GlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyrLysAsp 73
Db 979 GCTAGCTATGAAGATGGTAAG-----AAAGTGTTTGTATAAAGAAATTACATACAGTAAT 1032
QY 74 AsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeuArgAla 93
Db 1033 AACTCTCGTGTGCTAGCTAGGGCAAGTATTAATCTTGATGAAAATTGACCTTACGTCGG 1092
QY 94 GlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIleProAspThr 113
Db 1093 GGTATTGCTTACGATCAAGCGGCATCTCGTCATCACCGTAGTCGTCGAATTCAGATACC 1152
QY 114 AspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSerValAsp 133
Db 1153 GATCGCATTGGTATAGTTTAGGGCAACCTATATAATTCACGCCGAATTTATCTGTTGAT 1212
QY 134 ValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsnIle--- 152
Db 1213 CTTGGCTATGCTTACTTTAAAGGCAAAAAGTTCCACTTTAAAGAAAGTACAAAAGTTGCA 1272
QY 153 LysGlyLeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeuTyrGly 172
Db 1273 GATGGCTTGATCAACAAACCGCAATATACACTTCTCAAGCACACGCAATCTTTACGGT 1332
QY 173 LeuAsnLeuAsnTyrArgPhe 179
Db 1333 TTAACCTAAATTTATAGTTTC 1353
RESULT 9
AF260374 LOCUS
DEFINITION Haemophilus influenzae strain 9 outer membrane protein p1 (ompP1)
gene, complete cds.
ACCESSION AF260374
VERSION AF260374.1 GI:9716640
KEYWORDS

```

## SOURCE

Haemophilus influenzae.  
Haemophilus influenzae

Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
Haemophilus

## REFERENCE

1 (bases 1 to 1350)  
Boulduc, G.R., Bouchet, V., Jiang, R.Z., Geisselsoder, J.,  
Truong-Boulduc, C., Rice, P.A., Pelton, S.I. and Goldstein, R.  
Variability of outer membrane protein p1 and its evaluation as a  
vaccine candidate against experimental otitis media due to  
nontypeable Haemophilus influenzae: an unambiguous, multifaceted  
approach. Immun. 68 (8), 4505-4517 (2000)

## JOURNAL

MEDLINE  
20359342

## PUBMED

10899849

## REFERENCE

2 (bases 1 to 1350)

## AUTHORS

Bouchet, V.

## TITLE

Direct Submission

## JOURNAL

Submitted (25-APR-2000) Section of Molecular Genetics, The Maxwell  
Finland Laboratory for Infectious Diseases, Boston University  
School of Medicine, Boston Medical Center, 774 Albany Street,  
Boston, MA 02118, USA

## FEATURES

Location/Qualifiers

1..1350  
/organism="Haemophilus influenzae"  
/strain="9"  
/serotype="non-typable"  
/db\_xref="taxon:727"  
/country="USA"  
1..1350  
/gene="ompP1"  
1..1350  
/gene="ompP1"  
/note="OMP P1"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane protein p1"  
/protein\_id="AF97586.1"  
/db\_xref="GI:9716641"

/translation="MKKFNOSLLATAMLLAAGNAAAFOLAEVSTSGLRAYAGEAA  
IADNSVATFNPALMSLFKTAQFSTGCVYDPRINMGDVNASKATMNTKYGSAQ  
ERNVPCAFVFNLYFVAPNDKFAAGMNVNFKLSEYDDDDYDAGVGGKTDLSAINL  
NLNSGAVRTTEGLSLGLGVNAVYAKAQVERNAGIIVRSVKDSOAOATLKVPTLIP  
DLTSDKSVLSQDRAAWGFNAGVYQFNEANRIGLAYHSHKVIDDFDRTATSLG  
OKSIEAGKTNLTLPDYLELSPGHQTLDFAVHYSYKTHWSRLTLPDYASTEDGKK  
VFDELOYSNNRSVALGASYNLDEKLTLAGIAYDQAASRHRSAAIPDTRTWYSLG  
ATYKFTPNLSVDLGAYLKGKKVHEKVKVADGLITTNANTSOAHANLYGLNLNYSF"

BASE COUNT 417 a 235 c 293 g 405 t

ORIGIN

Alignment Scores:

Pred. No.: 1.02e-49 Length: 1350

Score: 562.50 Matches: 107

Percent Similarity: 76.70% Conservative: 28

Best Local Similarity: 60.80% Mismatches: 38

Query Match: 58.78% Indels: 3

DB: 1 Gaps: 2

US-09-995-493-52 (1-179) x AF260374 (1-1350)

QY 4 GlyValLeuGlyProTyrIleGlyLysGlySerLeuThrLeuLysLeuProAlaTyrTrp 23

Db 829 GGGGTTATCGGGCGGGG---AAAAAAGGTGATTAAACCCCTACATTCAGATTACTTA 885

QY 24 GluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLys 43

Db 886 GAACCTTCTGGTTTCCCATCAATTAACCAAACTTCGAGTATAGTATATAAATAT 945

QY 44 ThrGluTrpSerArgPheHisGlnLeuThrLysGlyLysTyrGlnAspGlySerGlyTyrGlu 63

Db 946 ACCCATTCGAGTCGTTTAAACAAATTAACCGCTAACTTCGAGATGGTAAA-----AAA 999

QY 64 AlaPheThrLysLysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThrThrTyr 83

```
Db 1000 GCGTTTGATAAAGATTACAATACAGTAAATAACTCTCGTTGTCATTAGGGCAAGTTAT 1059
Qy 84 SerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyAspLysAlaAlaSerLys 103
Db 1060 AATCTTTATGAAAAATTAACCTTACCTGCGGTATTCGTACGATCAAGCGCATCTGT 1119
Qy 104 ThrHisLeuSerAlaSerIleProAspThrAspArgMetTyrSerIleGlyAlaThr 123
Db 1120 CATCACCGTAGTCGCAATTCAGATACCGATCGCATCTGGTATAGTGGCAAC 1179
Qy 124 TyrLysPheThrProAsnLeuSerValAspValGlyPheAlaHisLeuArgGlyLys 143
Db 1180 TATAAATTCACCCCTAATTTATGCTGTGCTTACTTAAAGGCAACAA 1239
Qy 144 LysHisPheValGluThrGlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThr 163
Db 1240 GTTCACTTTTAAAGAAAGTACAAAGTAGTGGCCACATAAATAACAAACGCAANTTACACT 1299
Qy 164 ThrLysAlaThrAlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179
Db 1300 TCTCAAGCATGCAAACTTTACGGCTTAAACTTAAATATATAGTTTC 1347

RESULT 10
AF260353
LOCUS AF260353 1368 bp DNA linear BCT 07-AUG-2000
DEFINITION Haemophilus influenzae strain 161 outer membrane protein P1 (ompP1)
ACCESSION AF260353
VERSION AF260353.1 GI:9716598
KEYWORDS Haemophilus influenzae.
SOURCE Haemophilus influenzae.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
REFERENCE 1 (bases 1 to 1368)
AUTHORS Bolduc,G.R., Bouchet,V., Jiang,R.Z., Geisselsoder,J.,
TITLE Truong-Bolduc,Q.C., Rice,P.A., Pelton,S.I. and Goldstein,R.
VARIABILITY OF OUTER MEMBRANE PROTEIN P1 AND ITS EVALUATION AS A
VACCINE CANDIDATE AGAINST EXPERIMENTAL OTITIS MEDIA DUE TO
NONTYPEABLE HAEMOPHILUS INFLUENZAE: AN UNAMBIGUOUS, MULTIFACETED
APPROACH
JOURNAL Infect. Immun. 68 (8), 4505-4517 (2000)
MEDLINE 20359342
PUBMED 10899849
REFERENCE 2 (bases 1 to 1368)
AUTHORS Bouchet,V.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2000) Section of Molecular Genetics, The Maxwell
Finland Laboratory for Infectious Diseases, Boston University
School of Medicine, Boston Medical Center, 774 Albany Street,
Boston, MA 02118, USA
FEATURES
source
1. .1368
/organism="Haemophilus influenzae"
/strain="161"
/serotype="non-typable"
/db_xref="taxon:727"
/country="Finland"
1. .1368
/gene="ompP1"
1. .1368
/gene="ompP1"
/feature="OMP P1"
/codon_start=1
/transl_table=11
/product="outer membrane protein P1"
/protein_id="AAF97567.1"
/db_xref="GI:9716598"
/transl_table="MKRFNSILATAMLLAAGNAAAFQALAEVSTSGLRAYAGEAA
IADNAAVATNPALMSLFKTAQFGVYVDSIRNMNGDYDTASITATMSAKYGSAS
ERNVPGAFVFNLYFVAVNDKFKALGAMVNFGLKSEYDDSDYDAGIFGGKTDLTAIN
NLISGAYRVTEGLSLGIVNAVYAKAVERNAGLIANSVNDTOVOTALSAPPLKGL
DQNLPSKDKSVSVSLQDRAAGFGWAGVMTQFNEANRIGLAYHSKYVDIDFTDRATSV
```

```
BASE COUNT 424 a 235 c 293 g 416 t
ORIGIN
Alignment Scores:
Pred. No.: 1,49e-49 Length: 1368
Score: 561.00 Matches: 107
Percent Similarity: 79.65% Conservative: 30
Best Local Similarity: 62.21% Mismatches: 29
Query Match: 58.62% Indels: 6
DB: 1 Gaps: 3
US-09-995-493-52 (1-179) x AF260353 (1-1368)
Qy 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31
Db 856 AAGGTAATTTAAACCTTACATTCGACAGATTACTAGAACCTTCTGGGTTCATCAATTA 915
Qy 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51
Db 916 ACTGACAAATTTGCAGTCATTTATATATAATATACCATTCGAGTCGTTACAAAA 975
Qy 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71
Db 976 TTACACGCTAGCTTCGAAGATGGTAAG-----AAAGCTTTTGATAAAGAAATTACAATAC 1029
Qy 72 LysAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeu 91
Db 1030 AGTAATACTCTCGTGTTCATTCAGGCGCAAGTTATATCTTGATGAAATAATTAACCTTA 1089
Qy 92 ArgAlaGlyLeuAlaTyAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111
Db 1090 COTGCGGTATTGCTTACGATCAAGCTGCATCTCGTCATCACCCTAGTCGCAATTCCA 1149
Qy 112 AspThrAspArgMetTyrTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131
Db 1150 GATACCGATCGCAGCTTGTATAGTTTAGTGCAACCTATAAATTACGCCGCAATTTATCT 1209
Qy 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151
Db 1210 GTTGATCTTGCTATGCTTACTTAAAGCGCAAAAAGTTTCACTTTAAAGAAAGTAAACA 1269
Qy 152 Ile-----LysGlyLeuLeuLeu---ValGluAlaAspTyrThrThrLysAlaThr 167
Db 1270 ATAGTGACAAACGTCATTCGACATTGAATACAACTGCAAAATTATATCTTCAACGACAC 1329
Qy 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179
Db 1330 GCAAATCTTACGGTTTGAATTTAAATTTATAGTTTC 1365

RESULT 11
AF260354
LOCUS AF260354 1368 bp DNA linear BCT 07-AUG-2000
DEFINITION Haemophilus influenzae strain Ot-9 outer membrane protein P1
ACCESSION AF260354
VERSION AF260354.1 GI:9716600
KEYWORDS Haemophilus influenzae.
SOURCE Haemophilus influenzae.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
REFERENCE 1 (bases 1 to 1368)
AUTHORS Bolduc,G.R., Bouchet,V., Jiang,R.Z., Geisselsoder,J.,
TITLE Truong-Bolduc,Q.C., Rice,P.A., Pelton,S.I. and Goldstein,R.
VARIABILITY OF OUTER MEMBRANE PROTEIN P1 AND ITS EVALUATION AS A
VACCINE CANDIDATE AGAINST EXPERIMENTAL OTITIS MEDIA DUE TO
NONTYPEABLE HAEMOPHILUS INFLUENZAE: AN UNAMBIGUOUS, MULTIFACETED
APPROACH
JOURNAL Infect. Immun. 68 (8), 4505-4517 (2000)
```







gene

CDS  
1. .1368  
/gene="omp1"  
1. .1368  
/gene="omp1"  
/note="OMP P1"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane protein P1"  
/protein\_id="AAF97563.1"  
/db\_xref="GI:9716591"

BASE COUNT 428 a 235 c 290 g 415 t

ORIGIN

Alignment Scores:  
Pred. No.: 2,42e-49 Length: 1368  
Score: 559.00 Matches: 106  
Percent Similarity: 78.49% Conservative: 29  
Best Local Similarity: 61.63% Mismatches: 31  
Query Match: 58.41% Indels: 6  
DB: 1 Gaps: 2

US-09-995-493-52 (1-179) x AF260349 (1-1368)

QY 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31  
Db 856 AAAGGTAATTAACCCCTTACATGCCAGATTACTTAGAATCTTCGGTTCATCAATTA 915  
QY 32 ThrAspGlnTrpAlaLeuHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51  
Db 916 ACTGACAAATTCGCTGCTGCTAGGCGCAAGTATAATCTTCATGAAAAATTAACCTTA 975  
QY 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71  
Db 976 TTACACGCTAGCTTCAAGATGCTAG-----AAAGCTTTTCATTAAGAAATTAACATAC 1029  
QY 72 LysAspAsnSerArgPheAlaLeuGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeu 91  
Db 1030 AGTAATAACTCTGCTGCTAGGCGCAAGTATAATCTTCATGAAAAATTAACCTTA 1089  
QY 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111  
Db 1090 CGTCCGGGTATTGCTTACGATCAAGTCATCTCGTCATCACCGCTAGCTGCTGCAATTCCA 1149  
QY 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131  
Db 1150 GATACCGATCGCACTGGTATAGTTAGGTCACCACTTAAATTCACGCCGAATTTATCT 1209  
QY 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151  
Db 1210 GTTGATCTTGCTATGCTTACTTAAAGCGCAAAACCTTCACCTTAAAGAAAGTAAACA 1269  
QY 152 IleLysGly-----LeuLeuLeuValGluAlaAspTyrThrThrLysAlaThr 167  
Db 1270 ATAGGTGACAAAGCTACATGACATGAATACAACTGCAATATATACCTTCAAGCACAC 1329  
QY 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
Db 1330 GCAATCTTTACGGTTGAATTTAAATATATAGTTTC 1365

RESULT 15

AF260350

LOCUS

DEFINITION

Haemophilus influenzae strain 6107 outer membrane protein P1 (omp1) gene, complete cds.

ACCESSION

AF260350  
AF260350.1 GI:9716592

KEYWORDS

SOURCE  
ORGANISM  
Haemophilus influenzae.  
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
Haemophilus.

REFERENCE

1 (bases 1 to 1368)

AUTHORS

Bolduc G.R., Bouchet V., Jiang R.Z., Geisselsoder J.,  
Truong-Bolduc O.C., Rice, P.A., Pelton, S.I. and Goldstein, R.  
Variability of outer membrane protein P1 and its evaluation as a  
vaccine candidate against experimental otitis media due to  
nontypeable Haemophilus influenzae: an unambiguous, multifaceted  
approach

JOURNAL

Infect. Immun. 68 (8), 4505-4517 (2000)

MEDLINE

20359342

PUBMED

10899849

REFERENCE

2 (bases 1 to 1368)

AUTHORS

Bouchet, V.

TITLE

Direct Submission

JOURNAL

Submitted (25-APR-2000) Section of Molecular Genetics, The Maxwell  
Finland Laboratory for Infectious Diseases, Boston University  
School of Medicine, Boston Medical Center, 774 Albany Street,  
Boston, MA 02118, USA

FEATURES

Source

Location/Qualifiers

1. .1368

/organism="Haemophilus influenzae"

/strain="6107"

/serotype="b"

/db\_xref="taxon:727"

/country="United Kingdom"

1. .1368

/gene="omp1"

1. .1368

/gene="omp1"

/note="OMP P1"

/codon\_start=1

/transl\_table=11

/product="outer membrane protein P1"

/protein\_id="AAF97564.1"

/db\_xref="GI:9716593"

/translation="MKKFNQSLATAMLLAAGNAAAFQAEVSTSGIGRAYAGAA  
IADNASVATNPALMSLFKTAQFTSGGVYVDSRINMGDVTSIATTRKNSAKYGSAS  
ERNVPGAEVFNLYFVAPVNDKFAAGNANVFNGLKSEYDDSDYAGIFGGTDLTAI  
NLNSGAYRVTEGLSLGLGVNAVYAKAQRNAGIANSVNDTQVKTALSVLAPLKLGL  
DONLPSKDSVSVLODRAAGFGWAGVYQNEANRIGLAYHSKVDIDPTDRTATSV  
EANVLKEGKGKGLTLTPDYLELSGPHOLTDFAVHSYKYTHWSRLTKLHASFEDEK  
KAFDELQYNNRSVALGASYNLDEKLTLAGIAYDQASRRHRSRAIPDTRTWYSL  
GATYFTPNLSVDLGYALYKGGKVIHFKVETIGDKRTLTNTTANTYTSQAHANLYGLN  
LNYSF"

BASE COUNT 426 a 236 c 291 g 415 t

ORIGIN

Alignment Scores:

Pred. No.: 2,42e-49 Length: 1368

Score: 559.00 Matches: 106

Percent Similarity: 78.49% Conservative: 29

Best Local Similarity: 61.63% Mismatches: 31

Query Match: 58.41% Indels: 6

DB: 1 Gaps: 2

US-09-995-493-52 (1-179) x AF260350 (1-1368)

QY 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31

Db 856 AAAGGTAATTAACCCCTTACATGCCAGATTACTTAGAATCTTCGGTTCATCAATTA 915

QY 32 ThrAspGlnTrpAlaLeuHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51

Db 916 ACTGACAAATTCGCTGCTGCTAGGCGCAAGTATAATCTTCATGAAAAATTAACCTTA 975

QY 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71

Db 976 TTACACGCTAGCTTCAAGATGCTAG-----AAAGCTTTTCATTAAGAAATTAACATAC 1029

QY 72 LysAspAsnSerArgPheAlaLeuGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeu 91

Db 1030 AGTAATAACTCTGCTGCTAGGCGCAAGTATAATCTTCATGAAAAATTAACCTTA 1089

QY 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111

Db 1090 CGTCCGGGTATTGCTTACGATCAAGTCATCTCGTCATCACCGCTAGCTGCTGCAATTCCA 1149

QY 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131

Db 1150 GATACCGATCGCACTGGTATAGTTAGGTCACCACTTAAATTCACGCCGAATTTATCT 1209

QY 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151

Db 1210 GTTGATCTTGCTATGCTTACTTAAAGCGCAAAACCTTCACCTTAAAGAAAGTAAACA 1269

QY 152 IleLysGly-----LeuLeuLeuValGluAlaAspTyrThrThrLysAlaThr 167

Db 1270 ATAGGTGACAAAGCTACATGACATGAATACAACTGCAATATATACCTTCAAGCACAC 1329

```

Db 976 TTACACCGTAGCTTCGAGATGGTAAG-----AAAGCTTTTGATAAGAAATTACAATAC 1029
Qy 72 LysAspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeu 91
Db 1030 AGTAATAACTCTCGTGTGCAATGAGGCGCAAGTATATATCTTGATGAAGAAATTAACTTTA 1089
Qy 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111
Db 1090 CGTCGGGTATTGCTTACGATCAAGCTGCATCTCGTCATCACCGTAGTGTGCAATTTCCA 1149
Qy 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131
Db 1150 GATACCCATCGCACTCGGTATAGTTAGTGCACACCTATAAATTCACGCCGAATTTATCT 1209
Qy 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151
Db 1210 GTTGATCTTGGCTATGCTTACTTAAAGGCAAAAGTTTCACTTTAAGAAAGTAAAAACA 1269
Qy 152 IleLysGly-----LeuLeuLeuValGluAlaAspTyrThrThrLysAlaThr 167
Db 1270 ATAGGTGACAAACGTACATTTGACATTCGAATACAACTGCAAAATTTACTTCTCAAGCACAC 1329
Qy 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179
Db 1330 GCAAACTCTTACGGTTTGAATTTAAATATATAGTTTC 1365

RESULT 16
AF260351 1368 bp DNA linear BCT 07-AUG-2000
LOCUS
DEFINITION Haemophilus influenzae strain 7109, outer membrane protein PI
ACCESSION AF260351
VERSION AF260351.1 GI:9716594
KEYWORDS
SOURCE
ORGANISM
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
REFERENCE
1 (bases 1 to 1368)
BOLDUC,G.R., BOUCHET,V., JIANG,R.Z., GEISSELSODER,J.,
TRUONG-BOUDUC,O.C., RICE,P.A., PELTON,S.I. and GOLDSTEIN,R.
Variability of outer membrane protein PI and its evaluation as a
vaccine candidate against experimental otitis media due to
nontypeable Haemophilus influenzae: an unambiguous, multifaceted
approach
Infect. Immun. 68 (8), 4505-4517 (2000)
JOURNAL MEDLINE 20359342
PUBMED 10899849
REFERENCE
2 (bases 1 to 1368)
BOUCHET,V.
Direct Submission
Submitted (25-APR-2000) Section of Molecular Genetics, The Maxwell
Finland Laboratory for Infectious Diseases, Boston University
School of Medicine, Boston Medical Center, 774 Albany Street,
Boston, MA 02118, USA
FEATURES
source
1. .1368
/organism="Haemophilus influenzae"
/strain="7109"
/serotype="non-typable"
/db_xref="taxon:727"
/country="Sweden"
1. .1368
/gene="OmpP1"
1. .1368
/gene="OmpP1"
/feature="OMP PI"
/codon_start=1
/transl_table=11
/product="outer membrane protein PI"
/protein_id="AAF97565.1"
/db_xref="GI:9716595"
/translation="MKKFNQSLATAMLLAAGGANAATAFQAEVSTSGLGRAYAGEAA

```

```

BASE COUNT 427 a 235 c 292 g 414 t
ORIGIN
Alignment Scores:
Pred. No.: 2,42e-49 Length: 1368
Score: 559.00 Matches: 106
Percent Similarity: 78.49% Conservative: 29
Best Local Similarity: 61.63% Mismatches: 31
Query Match: 58.41% Indels: 6
DB: 1 Gaps: 2

US-09-995-493-52 (1-179) x AF260351 (1-1368)
Qy 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31
Db 856 AAAGTAATTTAAACCTTACATTCGCAAGTACTTAGAAGCTTCTGGGTTCCATCAATTA 915
Qy 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51
Db 916 ACTCACAAATTTGAGTGCATTTATATATATATACCCATTCGAGTCGTTTAAACAAA 975
Qy 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71
Db 976 TTACACCGCTAGCTTCGAGATGGTAAG-----AAAGCTTTTGATAAGAAATTACAATAC 1029
Qy 72 LysAspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeu 91
Db 1030 AGTAATAACTCTCGTGTGCAATTAGGCGCAAGTATATATCTTGATGAAGAAATTAACCTTA 1089
Qy 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111
Db 1090 CGTCGGGTATTGCTTACGATCAAGCTGCATCTCGTCATCACCGTAGTGTGCAATTTCCA 1149
Qy 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131
Db 1150 GATACCCATCGCACTCGGTATAGTTAGTGCACACCTATAAATTCACGCCGAATTTATCT 1209
Qy 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151
Db 1210 GTTGATCTTGGCTATGCTTACTTAAAGGCAAAAGTTTCACTTTAAGAAAGTAAAAACA 1269
Qy 152 IleLysGly-----LeuLeuLeuValGluAlaAspTyrThrThrLysAlaThr 167
Db 1270 ATAGGTGACAAACGTACATTCGACATTCGAATACAACTGCAAAATTTACTTCTCAAGCACAC 1329
Qy 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179
Db 1330 GCAAACTCTTACGGTTTGAATTTAAATATATAGTTTC 1365

RESULT 17
AF260352 1368 bp DNA linear BCT 07-AUG-2000
LOCUS
DEFINITION Haemophilus influenzae strain ATCC9795 outer membrane protein PI
ACCESSION AF260352
VERSION AF260352.1 GI:9716596
KEYWORDS
SOURCE
ORGANISM
Haemophilus influenzae.
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
REFERENCE
1 (bases 1 to 1368)
BOLDUC,G.R., BOUCHET,V., JIANG,R.Z., GEISSELSODER,J.,
TRUONG-BOUDUC,O.C., RICE,P.A., PELTON,S.I. and GOLDSTEIN,R.
Variability of outer membrane protein PI and its evaluation as a
vaccine candidate against experimental otitis media due to
nontypeable Haemophilus influenzae: an unambiguous, multifaceted
approach
Infect. Immun. 68 (8), 4505-4517 (2000)
JOURNAL MEDLINE 20359342
PUBMED 10899849
REFERENCE
2 (bases 1 to 1368)
BOUCHET,V.
Direct Submission
Submitted (25-APR-2000) Section of Molecular Genetics, The Maxwell
Finland Laboratory for Infectious Diseases, Boston University
School of Medicine, Boston Medical Center, 774 Albany Street,
Boston, MA 02118, USA
FEATURES
source
1. .1368
/organism="Haemophilus influenzae"
/strain="7109"
/serotype="non-typable"
/db_xref="taxon:727"
/country="Sweden"
1. .1368
/gene="OmpP1"
1. .1368
/gene="OmpP1"
/feature="OMP PI"
/codon_start=1
/transl_table=11
/product="outer membrane protein PI"
/protein_id="AAF97565.1"
/db_xref="GI:9716595"
/translation="MKKFNQSLATAMLLAAGGANAATAFQAEVSTSGLGRAYAGEAA

```

vaccine candidate against experimental otitis media due to nontypeable Haemophilus influenzae: an unambiguous, multifaceted approach  
 Infect. Immun. 68 (8), 4505-4517 (2000)

JOURNAL  
 MEDLINE  
 PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (25-APR-2000) Section of Molecular Genetics, The Maxwell

Finland Laboratory for Infectious Diseases, Boston University

School of Medicine, Boston Medical Center, 774 Albany Street,

Boston, MA 02118, USA

Location/Qualifiers

1. .1368

/organism="Haemophilus influenzae"

/strain="ATCC9795"

/serotype="b"

/db\_xref="taxon:727"

1. .1368

/gene="ompP1"

1. .1368

/gene="ompP1"

/note="OMP P1"

/codon\_start=1

/transl\_table=11

/product="outer membrane protein P1"

/protein\_id="AAF97566.1"

/db\_xref="GI:9716597"

/translation="MKFNQSLATAMLLAAGCANAFAQLAEVSTSGLGAYAGEAA

IADNASVATNPALMSLFKTAQFTSGGVYVDSIRNMNGDVTASIAITTKNSAKYGSAS

ERNVPGAFVPLNYFVAPVNDKFAAGNVAFLGKSEYDDSDYAGIFGCKTDLTAIN

LNLSGAYRVTEGLSLGLGVNAVYAKAQNAGIANSVNDQVKTALSVLAPPLRGL

DONLPSKDSVSLQDRAAGFGWAGVYQFNEANRIGLAYHSKVDIDFTDRTATSV

EANVKEGKGNLTLLPDYLELSCGHOLTDFKFAVHYSKYTHWSRLKHLASFEDGK

KAFDELQYSNNRVALGASYNLDEKLTLAGIYDQAAASRRHRSAAIPDTRTWYSL

GATYFTPNLSVDLGAYLKGKKVHFKEVKTIGDKRTLTLNTTANTYTSQAHANLYGLN

LNYSF"

BASE COUNT 425 a 234 c 293 g 416 t

ORIGIN

Alignment Scores:

Pred. No.: 2,42e-49 Length: 1368

Score: 559.00 Matches: 106

Percent Similarity: 78.49% Conservative: 29

Best Local Similarity: 61.63% Mismatches: 31

Query Match: 58.41% Indels: 6

DB: 1 Gaps: 2

US-09-995-493-52 (1-179) x AF260352 (1-1368)

QY 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31

DB 856 AAGGTAAATTTACCTTACATGCCAGATAGTACTTGAACCTTCTGGTCCATCAATTA 915

QY 32 ThrAspGlnTrpAlaLeuHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51

DB 916 ACTGACAAAATTCAGTCGATATAGTTATATAATATACCATCCAGTCGATCTTTAACAAA 975

QY 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71

DB 976 TTACACGCTAGCTTCCAGATGCTAAG-----AAGCTTTTGCATAAAGAAATTCACATAC 1029

QY 72 LysAspAsnSerArgPheAlaLeuGlyThrTyrSerLeuAsnAspAlaLeuThrLeu 91

DB 1030 AGTAATAACTCTCGTTCGATAGGGCCAGTATAATCTTGATGAAAATTAACCTTA 1089

QY 92 ArgAlaGlyLeuAlaTyrAspLysAlaLeuSerLysThrHisLeuSerAlaSerIlePro 111

DB 1090 CGTGGGGTATTCGTACGATCAAGCTGCATCTCGTCATCACCGTGTGCAATTCCTCA 1149

QY 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131

DB 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131

Db 1150 GATACCGATCGCATCTGGTATAGTTAGGTGCACACCTATAAAATTCACGCCGAATTTATCT 1209

QY 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151

Db 1210 GTTGATCTTGGCTATGCTTACTTAAAGGCAAAAAGTTTCACTTTTAAAGAAGTAAACA 1269

QY 152 IleLysGly-----LeuLeuLeuValGluAlaAspTyrThrThrLysAlaThr 167

Db 1270 ATAGGTGACAAACGATACATTGACATTGAATACAACTGCAAAATTAATCTTCTCAAGCAC 1329

QY 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179

Db 1330 GCAATCTTTACGCTTGAATTTAAATTAATAGTTTC 1365

RESULT 18

HEAOMPPIA

LOCUS

DEFINITION

Haemophilus influenzae outer membrane protein (OMPp1) gene,

complete CDS.

ACCESSION M63151 M27682

VERSION M63151.1 GI:148954

KEYWORDS

outer membrane protein.

SOURCE

Haemophilus influenzae (strain 1613) DNA.

ORGANISM

Haemophilus influenzae

Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

Haemophilus.

REFERENCE 1 (bases 1 to 1368)

AUTHORS

Munson, R. Jr., Grass, S., Einhorn, M., Bailey, C. and Newell, C.

TITLE

Comparative analysis of the structures of the outer membrane

protein P1 genes from major clones of Haemophilus influenzae type b

Infect. Immun. 57 (11), 3300-3305 (1989)

JOURNAL

MEDLINE

PUBMED

2572549

FEATURES

Location/Qualifiers

1. .1368

/organism="Haemophilus influenzae"

/strain="1613"

/db\_xref="taxon:727"

1. .1368

/gene="OMPp1"

1. .1368

/gene="OMPp1"

/codon\_start=1

/transl\_table=11

/product="outer membrane protein"

/protein\_id="AAA24991.1"

/db\_xref="GI:148955"

/translation="MKFNQSLATAMLLAAGCANAFAQLAEVSTSGLGAYAGEAA

IADNASVATNPALMSLFKTAQFTSGGVYVDSIRNMNGDVTASIAITTKNSAKYGSAS

ERNVPGAFVPLNYFVAPVNDKFAAGNVAFLGKSEYDDSDYAGIFGCKTDLTAIN

LNLSGAYRVTEGLSLGLGVNAVYAKAQNAGIANSVNDQVKTALSVLAPPLRGL

DONLPSKDSVSLQDRAAGFGWAGVYQFNEANRIGLAYHSKVDIDFTDRTATSV

EANVKEGKGNLTLLPDYLELSCGHOLTDFKFAVHYSKYTHWSRLKHLASFEDGK

KAFDELQYSNNRVALGASYNLDEKLTLAGIYDQAAASRRHRSAAIPDTRTWYSL

GATYFTPNLSVDLGAYLKGKKVHFKEVKTIGDKRTLTLNTTANTYTSQAHANLYGLN

LNYSF"

sig\_peptide 1. .66

mat\_peptide 67.1365

/gene="OMPp1"

/product="outer membrane protein"

BASE COUNT 427 a 234 c 291 g 416 t

ORIGIN

Alignment Scores:

Pred. No.: 2,42e-49 Length: 1368

Score: 559.00 Matches: 106

Percent Similarity: 78.49% Conservative: 29

Best Local Similarity: 61.63% Mismatches: 31

Query Match: 58.41% Indels: 6

DB: 1 Gaps: 2

US-09-995-493-52 (1-179) x HEAOMPPIA (1-1368)

```
Qy 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 856 AAGAGTAATTTAACCTTACATTCAGCATCTTACAGCACTTCTGCGTCCATCAATTA 915

Qy 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 916 ACTGACAAATTTGCGTTCATCTATAGTTATATAATACCCATCGGAGTTCGTTTAAACAAA 975

Qy 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 976 TTACACGCTAGCTTGAAGATGGTAAAG-----AAAGCTTTTGATAAGAATTACAATAC 1029

Qy 72 LysAspAsnSerArgPheAlaIleGlyThrTrpTyrSerLeuAsnAlaLeuThrLeu 91
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1030 AGTAATAACTCTCGTGTTCATCTAGGCGCAAGTTAATCTTGATGAAAAATTAACTTTA 1089

Qy 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1090 CGTGGGTATTCGTTACGATCAAGCTGCATCTCGTCATCACCGTAGTGCCTCAATTCCTCA 1149

Qy 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1150 GATACCGATCGACTTGTATAGTTAGTGCACCTATTAATTCACCGCGAATTTATCT 1209

Qy 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1210 GTTGATCTTGGCTATGCTTACTTAAAGCGCAAAAGTTCATCTTTAAAGAAAGTAAACAA 1269

Qy 152 IleLysGly-----LeuLeuValGluAlaAspTyrThrLysAlaThr 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1270 ATAGTGACAACAGCTACATTCGACATTCGATGAACTGCAATTAATCTTCAAGCACAC 1329

Qy 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1330 GCAATCTTTACGGTTTCAATTTAAATTAATATAGTTTC 1365

RESULT 19
AF260348 1368 bp DNA linear BCT 07-AUG-2000
LOCUS Haemophilus influenzae strain 7004 outer membrane protein p1
DEFINITION (ompP1) gene, complete cds.
ACCESSION AF260348
VERSION AF260348.1 GI:9716588
KEYWORDS Haemophilus influenzae.
SOURCE Haemophilus influenzae
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
REFERENCE 1 (bases 1 to 1368)
AUTHORS Bolduc,G.R., Bouchet,V., Jiang,R.Z., Geisselsoder,J.,
Truong-Bolduc,Q.C., Rice,P.A., Pelton,S.I. and Goldstein,R.
TITLE Variability of outer membrane protein p1 and its evaluation as a
vaccine candidate against experimental otitis media due to
nontypeable Haemophilus influenzae: an unambiguous, multifaceted
approach
JOURNAL Infect. Immun. 68 (8), 4505-4517 (2000)
MEDLINE 20359342
PUBMED 10899849
REFERENCE 2 (bases 1 to 1368)
AUTHORS Bouchet,V.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2000) Section of Molecular Genetics, The Maxwell
Finland Laboratory for Infectious Diseases, Boston University
School of Medicine, Boston Medical Center, 774 Albany Street,
Boston, MA 02118, USA
FEATURES Location/Qualifiers
source 1..1368
/organism="Haemophilus influenzae"
/strain="7004"
/serotype="b"
/db_xref="taxon:727"
/country="Netherlands"
```

```
gene 1..1368
/gene="ompP1"
CDS 1..1368
/gene="ompP1"
/note="OMP p1"
/codon_start=1
/transl_table=11
/product="outer membrane protein p1"
/db_xref="GI:9716588"
/translacion="MKKPNOSILATAMLLAAGGANAFAQLAEVSTSLGRAYACEAA
IADNASVATNPALMSLFKAQFSTGGYVDSRINMGDDVTASDATTMMNSAKYGSAS
ERNVPGAFVNPFLYFAPVNDKFAAGMNVFGLKSEYDDSDYDAGIFGKGKDTLTAIH
LNLGAYRTVEGLSLGLGVNAVAQVERNAGIANSVNDYQVKTALSVLAPPLKGL
DONLPSKDKSVVSLQDRAAGFGWAGVYQFNEANRIGLAYHSKVDIDFDRATSV
EANYIKGKKGNTLTLPDYLELSGHQLTDKFAVHYSYKTYTHSRITKTLHASFEDGK
KAPDKELQYSNNSRVALGASYNLDEKLTLAGIAYDQAAASHRHSRAITPDRTWYSL
GATYKFTPNLSVDLCYATLDEGKVHFREKVTIGDKRTLTLNTTANTYTSQAHANLYLN
LNYSF"
BASE COUNT 424 a 235 c 293 g 416 t
ORIGIN
Alignment Scores:
Pred. No.: 3.92e-49 Length: 1368
Score: 557.00 Matches: 106
Percent Similarity: 77.91% Conservativeness: 28
Best Local Similarity: 61.63% Mismatches: 32
Query Match: 58.20% Indels: 6
DB: 1 Gaps: 2
US-09-995-493-52 (1-179) x AF260348 (1-1368)
Qy 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 856 AAGAGTAATTTAACCTTACATTCAGCATCTTACAGCACTTCTGCGTCCATCAATTA 915

Qy 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 916 ACTGACAAATTTGCGTTCATCTATAGTTATATAATACCCATCGGAGTTCGTTTAAACAAA 975

Qy 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 976 TTACACGCTAGCTTGAAGATGGTAAAG-----AAAGCTTTTGATAAGAATTACAATAC 1029

Qy 72 LysAspAsnSerArgPheAlaIleGlyThrTrpTyrSerLeuAsnAlaLeuThrLeu 91
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1030 AGTAATAACTCTCGTGTTCATCTAGGCGCAAGTTAATCTTGATGAAAAATTAACTTTA 1089

Qy 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1090 CGTGGGTATTCGTTACGATCAAGCTGCATCTCGTCATCACCGTAGTGCCTCAATTCCTCA 1149

Qy 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1150 GATACCGATCGACTTGTATAGTTAGTGCACCTATTAATTCACCGCGAATTTATCT 1209

Qy 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1210 GTTGATCTTGGCTATGCTTACTTAAAGCGCAAAAGTTCATCTTTAAAGAAAGTAAACAA 1269

Qy 152 IleLysGly-----LeuLeuValGluAlaAspTyrThrLysAlaThr 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1270 ATAGTGACAACAGCTACATTCGACATTCGATGAACTGCAATTAATCTTCAAGCACAC 1329

Qy 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1330 GCAATCTTTACGGTTTCAATTTAAATTAATATAGTTTC 1365

RESULT 20
AF260373 1356 bp DNA linear BCT 07-AUG-2000
LOCUS Haemophilus influenzae strain 13-H-1157 outer membrane protein p1
DEFINITION (ompP1) gene, complete cds.
```

```

ACCESSION AF260373.1 GI:9716638
VERSION AF260373.1
KEYWORDS
SOURCE
ORGANISM
Haemophilus influenzae.
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
REFERENCE 1 (bases 1 to 1356)
AUTHORS Bolduc,G.R., Bouchet,V., Jiang,R.Z., Geisselsoder,J.,
Truong-Bolduc,Q.C., Rice,P.A., Pelton,S.I. and Goldstein,R.
TITLE Variability of outer membrane protein P1 and its evaluation as a
vaccine candidate against experimental otitis media due to
nontypeable Haemophilus influenzae: an unambiguous, multifaceted
approach
JOURNAL Infect. Immun. 68 (8), 4505-4517 (2000)
MEDLINE 20359342
PUBMED 10899849
REFERENCE 2 (bases 1 to 1356)
AUTHORS Bouchet,V.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2000) Section of Molecular Genetics, The Maxwell
Finland Laboratory for Infectious Diseases, Boston University
School of Medicine, Boston Medical Center, 774 Albany Street,
Boston, MA 02118, USA
FEATURES
source
Location/Qualifiers
1..1356
/organism="Haemophilus influenzae"
/strain="13-H-1157"
/serotype="non-typable"
/db_xref="taxon:727"
/country="USA"
1..1356
/gene="ompP1"
1..1356
/gene="ompP1"
/feature="OMP P1"
/codon_start=1
/transl_table=11
/product="outer membrane protein P1"
/protein_id="AAF97587.1"
/db_xref="GI:9716639"
/translation="MKKFNQSLIATAMLLAAGCANAAAFQAEVSTSGLGRAYAGEAA
IADNVAATNPALMSLFKTAQFSGGVYIDSRKNNNGVDVAYLOSQSKMFKYGSAS
QANVVPFAPVNPYLPVAPVNDKFGALGMVNVNAGIADSDQSVQPAQVAPQNKAT
GLNSGAYRVTGLSLGLGVAYANAQVNAQVNEANRIGLAVHSKVIDFTDRATSL
GYLTKSKSVSLQSSAAGFCWAGVNAQVNEANRIGLAVHSKVIDFTDRATSL
ACAIKAGKGGDITLTPDYLELSPHQLTKLAVHSYKYTHWSRLTKLNASFEDGKK
AFDKELQVSNRSRVALGASYNLAERKLTLAGTAYDQASRHSRAIPDTRTWISLG
F"
BASE COUNT 412 a 243 c 295 g 406 t
ORIGIN
Alignment Scores:
Pred. No.: 9,04e-49 Length: 1356
Score: 553.50 Matches: 105
Percent Similarity: 78.11% Conservative: 27
Best Local Similarity: 62.13% Mismatches: 34
Query Match: 57.84% Indels: 3
DB: 1 Gaps: 2
US-09-995-493-52 (1-179) x AF260373 (1-1356)
QY 12 LysGlySerLeuThrLeuLeuLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31
Db 853 AAGGTGATTTAAACCTTACATTCGCCAGATTACTTGAACCTTCTGCTTCCATCAATTA 912
QY 32 ThrAspGlnTrpAlaLeuHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51
Db 913 ACTGACAAACTTCGAGTCGCATATAGTTATATAATATACCATTCGGAGTCCTTTAACAAA 972
QY 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71

```

```

Db 973 TTAAACGCTAGCTTCGAAGATGCTAAG-----AAAGCTTTTGTATAAAGAAATTTACATATC 1026
QY 72 LysAspAsnSerArgPheAlaLeuGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeu 91
Db 1027 AGTAATAACTCTCGTTCGTTAGGGCAAGTTATAATCTTCTGCTGAAAAATTAACCTTA 1086
QY 92 ArgAlaGlyLeuAlaTyrAspTysAlaLeuSerLysThrHisLeuSerAlaSerIlePro 111
Db 1087 CGTGGGGTATTCTTACGATCAAGCGGCATCTCGTCATCAGCGTAGTCTGCAATTTCCA 1146
QY 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131
Db 1147 GATACCGATCGCACTTGGTATAGTTAGTGCAACCTATATAATTCACGCCGAATTTATCT 1206
QY 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysHisPheValGluThrGlnAsn 151
Db 1207 GTTGATCTTGGCTATCTTACTTTAAAGCGCAAAAAAGTTCACCTTTAAAGAAGTACAAAA 1266
QY 152 IleLys---GlyLeuLeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeu 170
Db 1267 GCTCAGGTGGCCACATAATAACACCGCAATTTACACTTCTCAAGCACACGCAAAATCTT 1326
QY 171 TyrGlyLeuAsnLeuAsnTyrArgPhe 179
Db 1327 TACGGCTTAAACTTAAATTTATAGTTTC 1353
RESULT 21
AF260375 Haemophilus influenzae strain 7424 outer membrane protein P1
LOCUS (ompP1) gene, complete cds.
ACCESSION AF260375
VERSION AF260375.1 GI:9716642
KEYWORDS
SOURCE Haemophilus influenzae.
ORGANISM Haemophilus influenzae.
REFERENCE 1 (bases 1 to 1353)
AUTHORS Bolduc,G.R., Bouchet,V., Jiang,R.Z., Geisselsoder,J.,
Truong-Bolduc,Q.C., Rice,P.A., Pelton,S.I. and Goldstein,R.
TITLE Variability of outer membrane protein P1 and its evaluation as a
vaccine candidate against experimental otitis media due to
nontypeable Haemophilus influenzae: an unambiguous, multifaceted
approach
JOURNAL Infect. Immun. 68 (8), 4505-4517 (2000)
MEDLINE 20359342
PUBMED 10899849
REFERENCE 2 (bases 1 to 1353)
AUTHORS Bouchet,V.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2000) Section of Molecular Genetics, The Maxwell
Finland Laboratory for Infectious Diseases, Boston University
School of Medicine, Boston Medical Center, 774 Albany Street,
Boston, MA 02118, USA
FEATURES
source
Location/Qualifiers
1..1353
/organism="Haemophilus influenzae"
/strain="7424"
/serotype="non-typable"
/db_xref="taxon:727"
/country="Kenya"
1..1353
/gene="ompP1"
1..1353
/gene="ompP1"
/feature="OMP P1"
/codon_start=1
/transl_table=11
/product="outer membrane protein P1"
/protein_id="AAF97589.1"
/db_xref="GI:9716643"
/translation="MKKFNQSLIATAMLLAAGCANAAAFQAEVSTSGLGRAYAGEAA

```

IADNASVATNPALMSLFKTAQSTGCGVYVDSRINMGDVAASIKKATAMGRTKYGSAS  
ARNVPGAFVNPILYFVAPVNDKFKALGAGMNVNFKLSEYDDSDYDAGIFGGKTDLSAIN  
LNLSGAYRVTEGLSLGLGVNAYAKAQRVNRAGTIADSIQDNOIQOALKAVDPQTKLH  
NYLTSKDKSVSLQDRAAGFNGWAGVYOFNEANRIGLAYHSHKVDIDFSDRATSLG  
AKDIVAGKTGDLTFLPDYLELSCFHQTLTKLAVHYSKYVTHWSRLTKLHASYENGEPK  
AFDKELOYSNNSRVALGASVNLKELTLRAGIAYDQASRHSRAALPDPDRWTYSIG  
ATYKFTPNLSVDLGAYLKGKKVHFKEEQAGGHITANTYTOAHANLYGLNLNYSF

BASE COUNT 416 a 239 c 297 g 401 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.15e-48 Length: 1353  
Score: 552.50 Matches: 105  
Percent Similarity: 77.65% Conservative: 27  
Best Local Similarity: 61.76% Mismatches: 35  
Query Match: 57.73% Indels: 3  
DB: 1 Gaps: 2

US-09-995-493-52 (1-179) x AF260375 (1-1353)

Qy 11 GlyLys---GlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHis 29  
Db 847 GGGAAACAGGTGATTTAACCTTTACATTCGCCAGATTACTTAGAACCTTCGTGTTCCAT 906  
Qy 30 GlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPhe 49  
Db 907 CAATTAACAGCAAACTTCGCGTGCATTATAGTTATAAATATACCCATTCGCGTTTCA 966  
Qy 50 LysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysGlu 69  
Db 967 ACAAAATACACGTAGCTAGTAATGAATGGTGAA-----AAGCGTTTGTATAAGAATTA 1020  
Qy 70 GluTyrLysAspAsnSerArgPheAlaIleGlyThrTrpTyrSerLeuAsnAspAlaLeu 89  
Db 1021 CAATACAGTAATAACTCTCGTGTGCATTAGGCGCAAGTTATATCTTTATGAAAATTA 1080  
Qy 90 ThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSer 109  
Db 1081 ACCTTAGCTGGGGTATTGGCTTAGCATCAAGCGCATCTCGTCATCAGCGTAGTGTGCA 1140  
Qy 110 IleProAspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsn 129  
Db 1141 ATTCAGATACCGATCGCACTTGGTATAGTTTGGGCAACCTATAAATTCACCCCGAAT 1200  
Qy 130 LeuSerValAspValGlyPheAlaHisLeuArgGlyLysLysHisPheValIleuThr 149  
Db 1201 TTATCTGTGATCTTGGCTATGCTTACTTAAAGGCAAAAGTTCACTTTTAAAGAAGAA 1260  
Qy 150 GlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsn 169  
Db 1261 CAACAACAGGTGGCCACATAATAACACCGCAAAATTAATCTTCAAGCACACGCAAT 1320  
Qy 170 LeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
Db 1321 CTTTACGGCTTAAACTTAAATTTATAGTTTC 1350

RESULT 22

AF260376

LOCUS

DEFINITION Haemophilus influenzae strain 4-H-1093 outer membrane protein P1

(ompP1) gene, complete cds.

ACCESSION AF260376

VERSION AF260376.1

KEYWORDS GI:9716644

SOURCE Haemophilus influenzae.

ORGANISM Haemophilus influenzae.

Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;

REFERENCE 1 (bases 1 to 1353)

AUTHORS Bolduc, G.R., Bouchet, V., Jiang, R.Z., Geisselsoder, J.,

Truong-Bolduc, O.C., Rice, P.A., Pelton, S.I., and Goldstein, R.

TITLE Variability of outer membrane protein P1 and its evaluation as a

vaccine candidate against experimental otitis media due to nontypeable *Haemophilus influenzae*: an unambiguous, multifaceted approach

Infect. Immun. 68 (8), 4505-4517 (2000)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

Location/Qualifiers

1. .1353

/organism="Haemophilus influenzae"

/strain="4-H-1093"

/serotype="non-typable"

/db\_xref="taxon:727"

/country="USA"

1. .1353

/gene="ompP1"

1. .1353

/gene="ompP1"

/note="OMP P1"

/codon\_start=1

/transl\_table=11

/product="outer membrane protein P1"

/protein\_id="AAF97590.1"

/db\_xref="GI:9716644"

/translation="MKRFNOSLLATAMLLAAGGANAANAFOLAEVSTSGLRVAGENA

IADNASVATNPALMSLFKTAQSTGCGVYIDSRINMGDVAASIKKATAMARTKYGSAS

QRNVPGAFVNPILYFVAPVNDKFKALGAGMNVNFKLSEYDDSDYDAGIFGGKTDLSAIN

LNLSGAYRVTEGLSLGLGVNAYAKAQRVNRAGTIADSIQDNOIQOALKAVDPQTKLH

EYLTSDKSVSLQDRAAGFNGWAGVYOFNEANRIGLAYHSHKVDIDFSDRATSLG

KKDIVAGKTGDLTFLPDYLELSCFHQTLTKLAVHYSKYVTHWSRLTKLHASYENGEPK

AFDKELOYSNNSRVALGASVNLKELTLRAGIAYDQASRHSRAALPDPDRWTYSIG

ATYKFTPNLSVDLGAYLKGKKVHFKEEQAGGHITANTYTOAHANLYGLNLNYSF

"

BASE COUNT 421 a 231 c 290 g 411 t  
ORIGIN

Alignment Scores:

Pred. No.: 1.15e-48 Length: 1353  
Score: 552.50 Matches: 105  
Percent Similarity: 78.24% Conservative: 28  
Best Local Similarity: 61.76% Mismatches: 34  
Query Match: 57.73% Indels: 3  
DB: 1 Gaps: 2

US-09-995-493-52 (1-179) x AF260376 (1-1353)

Qy 11 GlyLys---GlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHis 29  
Db 847 GGGAAACAGGTGATTTAACCTTTACATTCGCCAGATTACTTAGAACCTTCGTGTTCCAT 906  
Qy 30 GlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPhe 49  
Db 907 CAATTAACAGCAAACTTCGCGTGCATTATAGTTATAAATATACCCATTCGCGTTTCA 966  
Qy 50 LysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysGlu 69  
Db 967 ACAAAATACACGTAGCTAGTAATGAATGGTGAA-----AAGCGTTTGTATAAGAATTA 1020  
Qy 70 GluTyrLysAspAsnSerArgPheAlaIleGlyThrTrpTyrSerLeuAsnAspAlaLeu 89  
Db 1021 CAATACAGTAATAACTCTCGTGTGCATTAGGCGCAAGTTATAATCTTTATGAAAATTA 1080  
Qy 90 ThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSer 109  
Db 1081 ACCTTAGCTGGGGTATTGGCTTAGCATCAAGCGCATCTCGTCATCAGCGTAGTGTGCA 1140  
Qy 110 IleProAspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsn 129



```

|||||
Db 1141 ATTCCAGATACCGATCGCATTTGGTATAGTTAGGGCGCAACCTATATAATTTACCCCTAAT 1200
|||||
QY 130 LeuSerValAspValGlyPheAlaHisLeuArgGlyLysLysHisPheValGluThr 149
|||||
Db 1201 TTATCTGTTGATCTGGCTATGCTTACTTAAAGAGCAAAAGTTACCTTTAAAGANGTA 1260
|||||
QY 150 GlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThrLysAlaThrAlaAsn 169
|||||
Db 1261 AAAACAATAGGTGCCACACATAAACAACGCAAAATACACTTCTCAAGCACATGCAAT 1320
|||||
QY 170 LeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179
|||||
Db 1321 CTTTACGGTTTGAATTTAAATATATAGTTTC 1350
|||||

RESULT 23
LOCUS AF260367 1356 bp DNA linear BCT 07-AUG-2000
DEFINITION Haemophilus influenzae strain 4-H-1094 outer membrane protein P1
ACCESSION AF260367
VERSION AF260367.1 GI:9716626
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1..1356
/organism="Haemophilus influenzae"
/strain="4-H-1094"
/serotype="non-typable"
/db_xref="taxon:727"
/country="USA"
1..1356
/gene="omp1"
1..1356
/gene="omp1"
/feature="OMP P1"
/codon_start=1
/product="outer membrane protein P1"
/protein_id="AAF97581.1"
/db_xref="GI:9716627"
/translation="MKKEFQSLATAMLLAGGANAFAQLAEVSTSLGRAYAGEAA
IADNSVATPALMSLFTAQFSTGGYVDSRINMGDVKSSTKNLQMLAKYGSAS
QRNVVGFAPVNLXPVAPVNDKFCALGAMVNFGLKSEYDSYDAGVFGGKTDLAIN
LNLGAYRTVEGLSLGLGVNAVYAKAQVERNAGIIKESVELAKTAFTVAQDDKAAPT
YLTAKNTSVSLQDRAANGFCGNACVMQFENRIGLAYSFKVDIDFVDPATSLA
GVIGAKGKGLDTLTLPDYLELSGFHQLDKLAVHYSKYKTHNSRLTKLNASEDEKKA
FDKEQYNNRSLGASDLYELKTLTRAGTAYDQAASRHSKSAIPTDRTWYSLGA
TYKFPNLSVDGLAYLKGKVFHEKVTGTGKTLNATADYTSOAHANLYGLNLNS
P"

BASE COUNT 416 a 230 c 299 g 411 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 1..86e-48 Length: 1356
Score: 550.50 Matches: 109
Percent Similarity: 75.84% Conservative: 26
Best Local Similarity: 61.24% Mismatches: 38
Query Match: 57.52% Indels: 5
DB: 1 Gaps: 3

US-09-995-493-52 (1-179) x AF260367 (1-1356)

QY 4 GlyValLeuGlyProTyrIleGlyLysGlySerLeuThrLeuLysLeuProAlaTyrTrp 23
|||||
Db 829 GGGGTTATCGGGCGGGG---AAAAAAGGTGATTTAACCCCTTACATTTGCCAGATTTACTTA 885
|||||
QY 24 GluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLysTyr 43
|||||
Db 886 GAACCTTCTGGTTCATCATCACTTAAGTTCGCGTGCATTTATAGTTATTAATAT 945
|||||
QY 44 ThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGlu 63
|||||
Db 946 ACCATTGGAGTCGTTTAAACAAAATTTAAACGCCAGCTTCGAAGATGCTAAG-----AAA 999
|||||
QY 64 AlaPheThrLysLysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThrThrTyr 83
|||||
Db 1000 GCGTTTGTAAAGGAATTACAATACTCTCGTGTTCATTTAGGGCAAGTTAC 1059
|||||
QY 84 SerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLys 103
|||||
Db 1060 GATCTTTATGAAAAATTAACCTTACGTGAGGTATTTGCTTACGATCAAGCTGCTCGT 1119
|||||
QY 104 ThrHisLeuSerAlaSerIleProAspThrAspArgMetTrpTyrSerIleGlyAlaThr 123
|||||
Db 1120 CATCAGCGTAGTCTGCAATTCAGATCCGATCGCACCTGGGTATAGTTTAGGGCAACC 1179
|||||
QY 124 TyrLysPheThrProAsnLeuSerValaspValGlyPheAlaHisLeuArgGlyLysLys 143
|||||
Db 1180 TATAAATTCACGCCGAATTTATCTGTGTGCTTTGGGTATGCTTACTTAAAGGCAAAAA 1239
|||||
QY 144 LysHisPhe-----ValGluThrGlnAsnIleLysGlyLeuLeuValGluAlaAsp 161
|||||
Db 1240 GTTCACCTTAAAGAAAGTAAACACACAGGTAAACACAGTAATATTTGAATGCACTGCAGAT 1299
|||||
QY 162 TyrThrLysAlaThrAlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179
|||||
Db 1300 TATACTTCTCAAGCACATGCAATCTTACGGCTTAAACTTAAATTAATATAGTTTC 1353

RESULT 24
LOCUS AF260358 1371 bp DNA linear BCT 07-AUG-2000
DEFINITION Haemophilus influenzae strain 1071 outer membrane protein P1
ACCESSION AF260358
VERSION AF260358.1 GI:9716608
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

```

School of Medicine, Boston Medical Center, 774 Albany Street,  
Boston, MA 02118, USA

FEATURES  
source

1. .1371  
/organism="Haemophilus influenzae"  
/strain="1071"  
/serotype="non-typable"  
/db\_xref="taxon:727"  
/country="Finland"  
1. .1371  
/gene="ompP1"  
1. .1371  
/gene="ompP1"  
/note="OMP P1"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane protein P1"  
/protein\_id="AAF97572.1"  
/db\_xref="GI:9716609"  
/translation="MKFNQSLATAMLLAAGGANAFAQLAEVSTSGLRAYAGEAA  
IADNASVATNPALMSLFKAQSTGGVYVDSRINMGDVASSVVISNMVKATQDQSA  
SARNVPMCFCTKSLGKARNDFALCAGNMFGLKSEYDDSDYDAGIFGCKTDLGAI  
NLNSGARYRTEGLSLGLGVNAVYAKAQVERNAGTISDSVNDLSTGLAKPKPE  
LNKHLPSKDKSVSLQDRAAGFAGNAGVYQNEANRIGLAYHSKVDIDFTDRTATS  
LEAGVCAKGTGTLTLPDYLELSGFHQLTDKLVHYSYKTHWSRLTKLNASFEDG  
KKAFDKLOYSNRSRVALGASYNLYEKLTLAGIAYDQAAASRHQSRSAIIPDTRTWS  
LGATYKFTPNLSVDLGAYLKGKKVHKFKGTIADKSGPLTSTANTYTSQAHANLYGL  
NLNYSF"

BASE COUNT 411 a 241 c 300 g 419 t

ORIGIN

Alignment Scores:  
Pred. No.: 2,4e-48 Length: 1371  
Score: 549.50 Matches: 111  
Percent Similarity: 76.80% Conservative: 28  
Best Local Similarity: 61.33% Mismatches: 33  
Query Match: 57.42% Indels: 9  
DB: 1 Gaps: 5

US-09-995-493-52 (1-179) x AF260358 (1-1371)

Qy 4 GlyValLeuGlyProTyrIleGlyLys---GlySerLeuThrLeuLysLeuProAlaTyr 22  
|||||:|||||  
Db 838 GGGGTTATCGGG-----CGCGGGAACAGCTAATTAACCTTACATGCCAGATTAC 891  
|||||:|||||  
Qy 23 TrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLys 42  
|||||:|||||  
Db 892 TTAGAACTTTCTGGTTTCCATCAATTAACACAACTTCCGCTGCATTATAGTTATAAA 951  
|||||:|||||  
Qy 43 TyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyr 62  
|||||:|||||  
Db 952 TATACCCATTGGATGGCTGTTTAACAAAATTATACGCCAGCTTCGAAGATGGTAAG----- 1005  
|||||:|||||  
Qy 63 GluAlaPheThrLysLysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThr 82  
|||||:|||||  
Db 1006 AAACGGTTGATAAAGATTACATACAGTAATTAACCTCTCGTGTGCTAGGCGGCAAGT 1065  
|||||:|||||  
Qy 83 TyrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSer 102  
|||||:|||||  
Db 1066 TATAATCTTTATGAAATAAATTAACCTTACGTCGGGTATTGCTTACGATCAAGCTGCATCT 1125  
|||||:|||||  
Qy 103 LysThrHisLeuSerAlaSerIleProAspThrAspArgMetTrpTyrSerIleGlyAla 122  
|||||:|||||  
Db 1126 CGTCATCAGCGTAGTGTGCAATTCAGATACCCATCGCACTTCGCTATAGTTTAGGGGCA 1185  
|||||:|||||  
Qy 123 ThrTyrLysPheThrProAsnLeuSerValAspValGlyPheAlaHisLeuArgGlyLys 142  
|||||:|||||  
Db 1186 ACCTATAAATTAACCGCAATTTATCTGTGATCTTGGCTATGCTTACTTAAAGGCATA 1245  
|||||:|||||  
Qy 143 LysLysHisPheValGluThrGlnAsnIle-----LysGlyLeuLeuValGlu 159  
|||||:|||||  
Db 1246 AAAGTTCACTTTAAAGAGGAAAAACAACAGTACGACAAATCTGCAATTCGCTTCACTTCA 1305  
|||||:|||||

Qy 160 ---AlaAspTyrThrThrLysAlaThrAlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArg 178  
|||||:|||||  
Db 1306 ACTGCAATATACATCTTCTCAAGCACACGCAATCTTTACGGCTTAACCTTAATATAGT 1365  
|||||:|||||  
Qy 179 Phe 179  
|||  
Db 1366 TTC 1368  
|||  
RESULT 25  
AF260347  
LOCUS AF260347  
DEFINITION Haemophilus influenzae strain 7416 outer membrane protein P1  
ACCESSION AF260347  
VERSION AF260347.1 GI:9716586  
KEYWORDS  
SOURCE Haemophilus influenzae.  
ORGANISM Haemophilus influenzae  
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
Haemophilus.  
REFERENCE 1 (bases 1 to 1371)  
AUTHORS Bolduc,G.R., Bouchet,V., Jiang,R.Z., Geiselsoder,J.,  
Tuong-Bolduc,Q.C., Rice,P.A., Pelton,S.I. and Goldstein,R.  
TITLE Variability of outer membrane protein P1 and its evaluation as a  
vaccine candidate against experimental otitis media due to  
nontypeable Haemophilus influenzae: an unambiguous, multifaceted  
approach  
JOURNAL Infect. Immun. 68 (8), 4505-4517 (2000)  
MEDLINE 20359342  
PubMed 10899849  
REFERENCE 2 (bases 1 to 1371)  
AUTHORS Bouchet,V.  
TITLE Direct Submission  
JOURNAL Submitted (25-APR-2000) Section of Molecular Genetics, The Maxwell  
Finland Laboratory for Infectious Diseases, Boston University  
School of Medicine, Boston Medical Center, 774 Albany Street,  
Boston, MA 02118, USA  
FEATURES  
Location/Qualifiers  
1. .1371  
/organism="Haemophilus influenzae"  
/strain="7416"  
/serotype="a"  
/db\_xref="taxon:727"  
/country="Kenya"  
1. .1371  
/gene="ompP1"  
1. .1371  
/gene="ompP1"  
/note="OMP P1"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane protein P1"  
/protein\_id="AAF97561.1"  
/db\_xref="GI:9716587"  
/translation="MKFNQSLATAMLLAAGGANAFAQLAEVSTSGLRAYAGEAA  
IADNASVATNPALMSLFKAQSTGGVYIDSRINMGDVDSYDAGVFGKTDLTAI  
NLNSGARYRTEGLSLGLGVNAVYAKAQVERNAGIADSVKDNQVKTALTVQGEPLK  
LDKYLPSKDTSVSLQDRAAGFAGNAGVYQNEANRIGLAYHSKVDIDFTDRTATS  
VEANVIAKAGKGGDLTLPDYLELSGFHQLTDKLVHYSYKTHWSRLTKLNASFEDG  
KKAFDKLOYSNRSRVALGASYNLYEKLTLAGIAYDQAAASRHQSRSAIIPDTRTWS  
LGATYKFTPNLSVDLGAYLKGKKVHKFKGTIADKSGPLTNTTANTYTSQAHANLYGL  
NLNYSF"

BASE COUNT 426 a 230 c 292 g 423 t

ORIGIN

Alignment Scores:  
Pred. No.: 2,71e-48 Length: 1371  
Score: 549.00 Matches: 106  
Percent Similarity: 77.91% Conservative: 28  
Best Local Similarity: 61.63% Mismatches: 32  
Query Match: 57.37% Indels: 6  
DB: 1 Gaps: 3

US-09-995-493-52 (1-179) x AF260347 (1-1371)	
QY 12	LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31
Db 859	AAAGGTGATTAAACCTTACATTGCCAGATTACTTAGAATCTTCTGGTTCCATCAATTA 918
QY 32	ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51
Db 919	ACTGACAAACTTCTGCTGCAATTAATAATACCATTTGGAGTCTGTTTAAACAAA 978
QY 52	LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71
Db 979	TTAAACCGCTAGCTTCAAGATGCTAA-----AAAGCGTTTGTATGAAGAATTACAATC 1032
QY 72	LysAspAsnSerArgPheAlaIleGlyThrTyrTyrSerLeuAsnAspAlaLeuThrLeu 91
Db 1033	AGTAATAACTCTCGTTGCTTAGGGGCAAGTTATAATCTTTATGAGAANTTAACTTAA 1092
QY 92	ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111
Db 1093	CGTGGGGTATTCGTTACGATCAAGCGGCATCTCGTCATCAGCGTAGTGTGCAATTTCCA 1152
QY 112	AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131
Db 1153	GATACGATCGCATCTGGTATAGTTAGGGCAACCTATATAAATTCACGCCGAATCTATCT 1212
QY 132	ValAspValGlyPheAlaHisLeuArgGlyLysLysHisPheValGluThrGlnAsn 151
Db 1213	GTGTGATCTTGGCTATGCTTACTTAAAGGCAAAAAGTTTACACTTTTAAAGAACACG 1272
QY 152	Ile-----LysGlyLeu---LeuLeuValGluAlaAspTyrThrThrLysAlaThr 167
Db 1273	ATAGGTAAGAATCTCGCTTGGCTTGGCACTGACACTGCAAAATTTACACTTCTCAAGCAC 1332
QY 168	AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179
Db 1333	GCAATCTTTACGGCTTAACTTAAATTTATAGTTTC 1368
RESULT 26	
AF260360	
LOCUS	
DEFINITION	1371 bp DNA linear BCT 07-AUG-2000
ACCESSION	Haemophilus influenzae strain LO-127 outer membrane protein P1
VERSION	(comp1) gene, complete cds.
KEYWORDS	AF260360
SOURCE	AF260360.1 GI:9716612
ORGANISM	Haemophilus influenzae.
REFERENCE	Haemophilus influenzae.
AUTHORS	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
TITLE	Haemophilus.
JOURNAL	1 (bases 1 to 1371)
MEDLINE	Bolduc,G.R., Bouchet,V., Jiang,R.Z., Geisselsoder,J.,
PUBMED	Truong-Bolduc,Q.C., Rice,P.A., Pelton,S.I. and Goldstein,R.
REFERENCE	2 (bases 1 to 1371)
AUTHORS	Direct Submission
TITLE	Bouchet,V.
JOURNAL	Submitted (25-APR-2000) Section of Molecular Genetics, The Maxwell
FEATURES	Finland Laboratory for Infectious Diseases, Boston University
source	School of Medicine, Boston Medical Center, 774 Albany Street,
	Boston, MA 02118, USA
	Location/Qualifiers
	1. 1371
	/organism="Haemophilus influenzae"
	/strain="LO-127"
	/serotype="non-typable"
/db_xref="taxon:727"	
/country="Canada"	
1. 1371	
/gene="ompP1"	
1. 1371	
/note="OMP P1"	
/codon_start=1	
/transl_table=11	
/product="outer membrane protein P1"	
/protein_id="AAF97574.1"	
/db_xref="GI:9716613"	
/translation="MKKFNQSLATAMLLAAGNANAFAQLAEVSTSGIGRAYAGBA	
IADNSVATNPALMSLEKTAQSTGGVYVDSTINNGNVPSSVTNTSMKTTMDGSA	
SRNVVPGAFVNPVNFAPVNDKFKALGAGMNVFGLKSEYDDSDYAGVGGKTDLSAI	
NNLTCAYRVTEGLSLGLGNVAYAKAQVERNAGVIVDINDTQIKXALAALSEPKD	
PNYLTSDKDSVLSQDRAAGFGNAGVYQFNEANRIGLAYHSHKVDIDFTDRATYS	
VQANVTKAGKGDLTLTLDYLELGSFQHLTKLAVHYSIKYTHRSRLTKLHASPEDG	
KFAFDKELQYSNNRSVALGANSYLKELTKLRAGIAYDQAASRHRSAATPDRTWYS	
LGATYKFTPNLSVDLGYAYLKGKVKHFEKVNNTIGRESGLPLTTTANYTTSQAHANLYGL	
NLNYSF"	
BASE COUNT	422 a 244 c 290 g 414 t 1 others
ORIGIN	
Alignment Scores:	
Pred. No.:	2,71e-48 Length: 1371
Score:	549.00 Matches: 107
Percent Similarity:	76.74% Conservative: 25
Best Local Similarity:	62.21% Mismatches: 34
Query Match:	57.37% Indels: 6
DB:	1 Gaps: 3
US-09-995-493-52 (1-179) x AF260360 (1-1371)	
QY 12	LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31
Db 859	AAAGGTGATTAAACCTTACATTGCCAGATTACTTAGAATCTTCTGGTTTCCATCAATTA 918
QY 32	ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51
Db 919	ACTGACAAACTTCCCGTGCATTTATAGTTATAAATATACCATTTGGAGTCTGTTTAAACAAA 978
QY 52	LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71
Db 979	TTACATGCCAGCTTCGAAGATGCTAAA-----AAAGCGTTTGTATGAAGAATTACAATC 1032
QY 72	LysAspAsnSerArgPheAlaIleGlyThrTyrTyrSerLeuAsnAspAlaLeuThrLeu 91
Db 1033	AGTAATAACTCTCGTTGCTTAGGGGCAAGTTATAATCTTTATGAGAANTTAACTTAA 1092
QY 92	ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111
Db 1093	CGTGGGGTATTCGTTACGATCAAGCGGCATCTCGTCATCAGCGTAGTGTGCAATTTCCA 1152
QY 112	AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131
Db 1153	GATACGATCGCATCTGGTATAGTTAGGGCAACCTATATAAATTCACGCCGAATCTATCT 1212
QY 132	ValAspValGlyPheAlaHisLeuArgGlyLysLysHisPheValGluThrGlnAsn 151
Db 1213	GTGTGATCTTGGCTATGCTTACTTAAAGGCAAAAAGTTTACACTTTTAAAGAACACG 1272
QY 152	Ile-----LysGlyLeuLeuLeu---ValGluAlaAspTyrThrThrLysAlaThr 167
Db 1273	ATAGGTGACGAAGCTTCACTTACATTTGAATACAACTGCAAAATTTATCTTCTCAAGCACAT 1332
QY 168	AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179
Db 1333	GCAATCTTTACGGTTTAACTTAAATTTATAGTTTC 1368

DEFINITION Sequence 5 from patent US 6355450.  
ACCESSION AR199524  
VERSION AR199524.1 GI:20249598  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1381)  
AUTHORS Fleischmann, R.D., Adams, M.D., White, O., Smith, H.O. and Venter, J. Craig.  
TITLE Computer readable genomic sequence of Haemophilus influenzae Rd, fragments thereof, and uses thereof  
JOURNAL Patent: US 6355450-A 5 12-MAR-2002;  
FEATURES Location/Qualifiers  
1..1381  
source /organism="unknown"  
BASE COUNT 424 a 237 c 297 g 423 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 2,74e-48 Length: 1381  
Score: 549.00 Matches: 106  
Percent Similarity: 77.91% Conservative: 28  
Best Local Similarity: 61.53% Mismatches: 32  
Query Match: 57.37% Indels: 6  
DB: Gaps: 3  
US-09-995-493-52 (1-179) x AR199524 (1-1381)  
Qy 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31  
Db 872 AAAGGTGATTTAAACCTTTACATTCGCAGATTACTTAGAACCTTCTGGTTTCCCAATTA 931  
Qy 32 ThrAspGlnTTPAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51  
Db 932 ACTGACAACTTCGTCGATATAGTTATATAATATACCCATTGGAGTCGTTTACAAAA 991  
Qy 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71  
Db 992 TTAACACGCTAGCTTGAAGATGGTAAA-----AAAGCTTTTGATAAAGAATTACAATAC 1045  
Qy 72 LysAspAsnSerArgPheAlaIleGlyThrTyrTrpSerLeuAsnAspAlaLeuThrLeu 91  
Db 1046 AGTAATAACTCTCGTTGGTTCATAGGGCAAGTTATAATCTTTGATAAAGAATTGACCTTA 1105  
Qy 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111  
Db 1106 CGTGGGGTATTGCTTACGATCAGCTGCATCTCGTCATCAGCGTAGTGTCAATTCCA 1165  
Qy 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131  
Db 1166 GATACCGATCGCAGCTGGTATAGTTTAGGGCAACCTATAAATTCACGCCGAATTTATCT 1225  
Qy 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151  
Db 1226 GTTGATCTTGGTATGCTTACTTAAGAGCAAAAAGTTCACTTTAAAGAGCTAAACA 1285  
Qy 152 Ile-----LysGlyLeuLeuLeu---ValGluAlaAspTyrThrLysAlaThr 167  
Db 1286 ATAGTGACGAACGTTCACTTACATTGAATACAACCTGCAAAATTATCTCTCAAGCACAT 1345  
Qy 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
Db 1346 GCAAACTTTACGGTTTAACTTAAATTAATATAGTTTC 1381  
RESULT 28  
AR085991  
LOCUS AR085991 1560 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 1 from patent US 5985288.  
ACCESSION AR085991  
VERSION AR085991.1 GI:10012757  
KEYWORDS  
SOURCE Unknown.

ORGANISM Unknown.  
Unclassified.  
REFERENCE 1 (bases 1 to 1560)  
AUTHORS Munson, R.S. Jr., Grass, S., Chong, P., Yang, Y.-P., Fahim, R., Charles, Sja, D. Yan., McVerry, P. and Klein, M.  
TITLE Outer membrane protein P1 and peptides of Haemophilus influenzae type B  
JOURNAL Patent: US 5985288-A 1 16-NOV-1999;  
FEATURES Location/Qualifiers  
1..1560  
source /organism="unknown"  
BASE COUNT 487 a 271 c 321 g 481 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3,15e-48 Length: 1560  
Score: 549.00 Matches: 105  
Percent Similarity: 76.74% Conservative: 27  
Best Local Similarity: 61.05% Mismatches: 34  
Query Match: 57.37% Indels: 6  
DB: Gaps: 2  
US-09-995-493-52 (1-179) x AR085991 (1-1560)  
Qy 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31  
Db 922 AAAGGTAATTTAACCTTTACATTCGCAGATTACTTAGAACCTTCTGGTTTCCCAATTA 981  
Qy 32 ThrAspGlnTTPAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51  
Db 982 ACTGACAACTTCGTCGATATAGTTATATAATATACCCATTGGAGTCGTTTACAAAA 1041  
Qy 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71  
Db 1042 TTACATGCCAGCTTCGAAGATGGTAAA-----AAAGCTTTTGATAAAGAATTACAATAC 1095  
Qy 72 LysAspAsnSerArgPheAlaIleGlyThrTyrTrpSerLeuAsnAspAlaLeuThrLeu 91  
Db 1096 AGTAATAACTCTCGTTGGTTCATAGGGCAAGTTATAATCTTTGATAAAGAATTGACCTTA 1155  
Qy 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111  
Db 1156 CGTGGGGTATTGCTTACGATCAGCGGCATCTCGTCATCAGCGTAGTGTCAATTCCA 1215  
Qy 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131  
Db 1216 GATACCGATCGCAGCTGGTATAGTTTAGTGCAACCTATAAATTCACGCCGAATTTATCT 1275  
Qy 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151  
Db 1276 GTTGATCTTGGTATGCTTACTTAAGAGCAAAAAGTTCACTTTAAAGAGCTAAACA 1335  
Qy 152 IleLysGly-----LeuLeuLeuValGluAlaAspTyrThrLysAlaThr 167  
Db 1336 ATAGTGACAAACGTCATCTGACATTGAATACAACCTGCAAAATTATCTCTCAAGCACAC 1395  
Qy 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
Db 1396 GCAAACTTTACGGTTTGAATTTAAATTAATATAGTTTC 1431  
RESULT 29  
HEAOMPP1  
LOCUS HEAOMPP1 1598 bp DNA linear BCT 26-APR-1993  
DEFINITION Haemophilus influenzae type b, outer membrane protein (omp) P1  
gene, complete cds.  
ACCESSION J03381  
VERSION J03381.1 GI:148952  
KEYWORDS outer membrane protein.  
SOURCE H. influenzae type b DNA, clone pRSM188.  
ORGANISM Haemophilus influenzae  
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
Haemophilus.  
REFERENCE 1 (bases 1 to 1598)

Db 1353 ATAGGTGACAACGTCACATTGACATTGAATACAACCTGCAAAATTACTTCTCAAGCACAC 1412

Qy 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
|||||  
1413 GCAAATCTTTACGGTTTGAATTTAAATATATAGTTTC 1448

Db

## RESULT 30

	AR089267	LOCUS	AR089267	1598 bp	DNA	linear	PAT# 07-SEP-2000
--	----------	-------	----------	---------	-----	--------	------------------

LOCUS AR069267 1596 bp DNA  
DEFINITION Sequence 26 from patent US 5994066.  
ACCESSION AR069267  
FAI 0/-SEP-2000

AR089267  
AR089267.1 GI:10016024

**KEYWORDS** . Unknown,  
**SOURCE** Unknown,  
**OPACITEM** Unknown

ORGANISM	REFERENCE
Unknown.	1, 4, 1500
Unclassified.	1, 4, 1500

REFERENCE  
I (bases 1 to 1598)  
Bergeron, M.G., Picard, F.J., Ouellette, M. and Roy, P.H.  
AUTHORS

**TITLE** Species-specific and universal DNA probes and amplification primers to rapidly detect and identify common bacterial pathogens and

associated antibiotic resistance genes from clinical specimens for routine diagnosis in microbiology laboratories

JOURNAL  
FEATURES  
Patent: US 5994066-A 26 30-NOV-1999;  
Location/Qualifiers

```
source
1. .1598
/organism="unknown"
```

BASE COUNT	500 a	278 c	329 g	491 t
ORIGIN				

Alignment Scores:

Length:	1598
Matches:	105
Score:	549.00
Pred. No.:	3.24e-48

Score:	76.74%	Conservative:	27
Percent Similarity:	76.74%	Mismatches:	105
Best Local Similarity:	61.05%		

Query Match:	57.37%	Indels:	6	Mismatches:	34
Best Local Similarity:	61.03%				
Na.	na.	Calls:	6		

DB:	0	z
115-00-085-403-52	1-170	1-1508

US-09-993-493-52 (1-1/9) X AR08926/ (1-1398)

09 LysGlySerLeuThrLeuLysLeuProAlaIyrTrpGluLeuSerGlyPheHisGlnLeu 31  
12  
030  
030

DB 939 ARAGGTAATTTAAAGCTTTACATTGCCAGATTACTTAGAAGCTTTCTGGTTTCCATCAATTA 998

Qy 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51  
|||||:: ||||:||||| ||||| ||||| |||||  
|||:: ||||:||||| ||||| ||||| |||||

Db 999 ACTGACAACTTGCAGTGCATTATACTTATAAATATACCCATTGGAGTCGTTTACAAA 1058

Qy 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaphethrLysLysGluCluTyr 71

Db 1059 TTACATGCCAGCTTCGAAGATGGTAAA-----AAAGCTTTTGATAGAATTTACAATAC 1112

Qy 72 LysAspAsnSerArgpHeAlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeu 91  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 1113 AGTAAATACTCTCGTGTTCGATTAGGGGCAAGTTATAATCTTTATGAAAAATTGACCTTA 1172

Qy . 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111

Db 1173 CGTGGGGTATTGGTTACGATCAAGGGGCATCTCGTCATCACCGTAGTCTGCAATTCCA 1232

Qy 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131  
|||||

Db 1233 GATACCGATCGCACTTGGTATAGTTAGGTGCAACCTATAAAATTCACGCCGAATTTATCT 1292

QY 132 valaspvalglyphealaHisLeuArgGlyLysLysHisPheValGluThrClnAsn 151

Db 1293 GTTGATCTTGGCTATGCTTAAAGGCCAAAAGTTCACATTAAAGAGTAAACCA 1352

Qy 152 IleLysGly.....LeuLeuValGluAlaAspTyrThrThrLysAlaThr 167

Db 1353 ATAGGTGACAAACGGTACATTGACATTGAATACAACTGCAAAATATATACTTCTCAAGCAC 1412

QY 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179

—

```
/usr_x86_64-01.15/3504
/translation="MDIFFVISGLITGIIITEIQONSFLKQFYTRRIKRIYPAFIT
VMALVSFIASAIIFYNDENKLRKTTIELAIAFLSNFYLGLTQGYFDLSANENPVLHIWS
```

```

LAVEGQYLIFPLILLILAYKKREVKVLFIITLILFFILLATSFVSANFYKEVLHQPNI
IYLSNLRPELLVGLSLAIYHNLKNKQVLSQVNNILAILSTLLFLSCFLMNNIA
FIPGTLILPICIPTALIIHTTSQNNIR"
complement(1195...2286)
/gene="HI0393"
CDS
complement(1195...2286)
/gene="HI0393"
/note="similar to SP:P31216 GB:U02423 GB:X61941 PID:311421
PID:416043 percent identity: 81.54; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical GTP-binding protein"
/protein_id="AAC22052.1"
/db_xref="GI:1573365"
/translacion="MGFKGIVGLPNVKGSTLFLNALTRAGIEAANYPFCTIETPNVGV
PMPDRLDALAEIVAPERILPTMEFVDIAGIVAGASGEGIGKXFLANIRETDAIGH
VYRCPENDIIVHAGKIDPLDDIDINTELALADLDSCERAQOROKRAGKGDKEAKF
ELSVMEKILVLNAGMIRSVGLDKEELQAIKSNFLTLKPTMYIANVNEDEGFENPNY
LDRVREIAAKEGAVVPCAAIESEIAELDDKEVEFFLQDLGIEEPGLNRVIRAGYAL
NLQTYFTAGVREVRAMTVSVGATAPKAAAVIHTDFEKFIRAEVIAYEDFIQFNGEN
GAKGAKWRLECKDYIVQDGVVHFRFNV"
complement(2335...2919)
/gene="HI0394"
CDS
complement(2335...2919)
/gene="HI0394"
/note="similar to SP:P23932 GB:X61941 PID:581202 GB:U000096
PID:1651597 percent identity: 64.21; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptidyl-trna hydrolase (pth)"
/protein_id="AAC22053.1"
/db_xref="GI:1573366"
/translacion="NSEIKLIVGLNPGDKYTDTRNAGEWLIERLARFNVLNAPES
KFEQTAPLVNGKEVRLVPTFMNLSCKAVGALASYRIKPEELILVHDELDPG
TAKLKQGGHGGHGLKDIYVQGLGNNNNFYRLIGIGHPGHRDLVAGYVLRNRPSPADR
DALEKVLDEATDCVEMIFRDMGVKATNRNLNPKFI"
3086...3394
/gene="HI0395"
CDS
3086...3394
/gene="HI0395"
/note="similar to PID:1033114 SP:P52119 GB:U000096
PID:1788971 percent identity: 55.32; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC22054.1"
/db_xref="GI:1573377"
/translacion="MNQINIEIAYAFPERYLYKSFQVDEGITVOTAITQSGILSOPPE
IDLSTNKGIFSRPIKLTDLVLRGEGRIEYRPLDAPKEIRKRAAEQAANKDEKGA
"
3397...4611
/gene="HI0396"
CDS
3397...4611
/gene="HI0396"
/note="similar to SP:P27431 GB:U000096 PID:1787373 percent
identity: 32.58; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC22055.1"
/db_xref="GI:1573367"
/translacion="WTFALSSVDFCLPEHITPEIFLRDYQKKPLVIRNGLPELVGQFE
PDQITELAQNEVDVTLARLVKTFSDDDMKVFFSPKSEKDFOLPEKMSVLVONLEQWSE
LAQNMKNKGFIFQWRDDIMVSYAPKGVSGKHGYDEYDLVQCYGHRMRQVQKWCDA
STFEPKPNQIRTFDDMGELVIDEVMNPGDILYIPARMAHYGVAEDDCLTFSTFGLRYPN
LNSLIDGISKGFCHODPLRPLRLOSORTKGLADENIOMKOLLKDLKLAH
SEAFDTLFQAAVASAVSSRRYELLYSDCMCDPDEVRISILEEDGAFILSDONCKLLYTE
NPLRIYANGEWELNIIIESEVLKRLSDGESLDWAFSLDLANKTEDPETSMDLLLDSI
CNWVDDGWALIE"

```

```

gene
4613...5932
/gene="HI0397"
CDS
4613...5932
/gene="HI0397"
/note="similar to GB:J02599 SP:P04994 PID:148275 GB:U000096
PID:1788856 percent identity: 57.69; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="exodeoxyribonuclease VII, large subunit (xseA)"
/protein_id="AAC22056.1"
/db_xref="GI:1573368"
/translacion="MSDNIYSVQLNSAARQMLEGNFCQIWLITGEISNTQPVSGHYY
LGLKDNDAQVRCAMFRMKNLRAVFPQNGQVILVRANVSYITSPRGDYQILIDSMHPAG
EFLQQQFALKMLAAEGLFAQNLKTLPHFSKAVGIITSTGTALODILHLIARRD
PSLKVVIYPTAVQGEATAEIVQMTELANARQEVDLIVRGSGSLEDLWCFNEEVA
RAIFRSTPLIIISAVGHETDVTIADFADLRAPTPSAAAEVLVSRNODQLAQQLRHQOOR
LMAFDRLETRKSQRKLQALRLQNHQPNQQLRAQAKNEQLTHRLQLAAILRQFENTQ
QFTALSVRLKQNPUPYRIQRIQOQGLEUKLRNFCVNRQVTERQNKLATLQCKLDGL
SPKLVLARGYSIAENPQGAIVSKVDVNGQDFITTOVADGKIVSKYL"
6109...6762
/gene="HI0398"
CDS
6109...6762
/gene="HI0398"
/note="similar to SP:P36651 PID:882564 GB:U000096
PID:1789412 percent identity: 54.00; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC22057.1"
/db_xref="GI:1573369"
/translacion="MQFWRKQMSIEQHFQQDIEILGEQTLVEGFTLKRIOFKKHLF
AGQSGVTVTRELLIKGASAVIADPKEDSVILVEQVRIGAAVHPESHSPMLLELIA
GMVEKGEKEDVALRESEEGAGTQVKNLTHCLSVDPGSIVERIHLFAGEVDSAAK
GTHGLAENEDIKVHVVKREQAYQWCEKIDNGIAVIGLQNLQNLVLAQLOQSWKRS"
6772...7596
/gene="HI0399"
CDS
6772...7596
/gene="HI0399"
/note="similar to GB:D16557 SP:P36650 PID:453396
PID:882562 GB:U000096 percent identity: 53.26; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="lacZ expression regulator (icc)"
/protein_id="AAC22058.1"
/db_xref="GI:1573370"
/translacion="MANTFYQAEKPVIKLLQITDPHLFKDESAELLGVNTQASPAQV
LKEIQENNEFDVILATGDLVDSSDEGYIRFVEMMKPFNKPVFIIPGNHDFQPKMAE
FLNQPPMNAKHLGLGHWQALLDSQYGVPHGQLSQHLDLKLTGKNPRTYLLV
VLHHLPLTNSAWLDOHNLNRNLSHELAEVLAPITNKAIIYGHIGHQVNSWNGYQVMA
TPATCIQKPKDCQYFSLDTPLOPCWREIELHSDGSIIRTEVKRIQQAEEFFPMQEEGY"
complement(7694...7981)

```

## Alignment Scores:

Pred. No.:	3-76e-47	Length:	13070
Score:	549.00	Matches:	106
Percent Similarity:	77.91%	Conservative:	28
Best Local Similarity:	61.63%	Mismatches:	32
Query Match:	57.3%	Indels:	6
DB:	1	Gaps:	3

US-09-995-493-52 (1-179) x U32723 (1-13070)

Qy	12	LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu	31
Db	9055	AAAGGTATTAAACCCCTTACCATTCGACGATTTAGCAACATTTCTGGTTTTCCATCAATTA	9114
Qy	32	ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu	51
Db	9115	ACTGACAAACTTGCTGTGTCATATAGTATATAATATACCATTCGGAGTCGTTTAAACAAA	9174







## ORIGIN

## Alignment Scores:

Pred. No.: 4,43e-48 Length: 1380  
 Score: 547.00 Matches: 106  
 Percent Similarity: 77.33% Conservative: 27  
 Best Local Similarity: 61.63% Mismatches: 33  
 Query Match: 57.16% Indels: 6  
 DB: 1 Gaps: 3

US-09-995-493-52 (1-179) x AF260344 (1-1380)

QY 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31  
 DB 868 AAGGTGATTAAACCTTACATTCAGCAGATTACTTAGAACCTTTCGGTTCCATTA 927  
 QY 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTyrSerArgPheLysGlu 51  
 DB 928 ACTGACAACTTGGCGTGCATTATAGTATATAATATATACCCATTGGAGTCGTTTACAAAA 987  
 QY 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluTyr 71  
 DB 988 TTAACCGTAGCTTCGAAGTGGTAAA-----AAAGCGTTTGATAAGAAATTACAATAC 1041  
 QY 72 LysAspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeu 91  
 DB 1042 AGTAATACTCTCGATTGGCATTAGGGGCAAGTTATATCTTTATGAAAACTCACCTTA 1101  
 QY 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLysLeuSerAlaSerIlePro 111  
 DB 1102 CGTGGGGTATTGCTTAGCATCAACGGCATCTCGTCATCAGCGTAGTGTGCAATTCCA 1161  
 QY 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131  
 DB 1162 GATACCGATCGCACTTGGTATAGTTAGGGCAACCTATAAATTCACGCCGAATTATCT 1221  
 QY 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151  
 DB 1222 GTTGATCTTGGCTATGCTTACTTAAAGGCAAAAAAGTTTCACTTTAAAGAAATAAACA 1281  
 QY 152 Ile-----LysGlyLeuLeuLeu---ValGluAlaAspTyrThrThrLysAlaThr 167  
 DB 1282 ATAGTGACGAAGCTTCACTTGCATTGAATGAACTGCAAAATTATATCTTCAACGACAT 1341  
 QY 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
 DB 1342 GCAAATCTTACGGCTTAAACTTAAATATATAGTTTC 1377

RESULT 36 AF260345 1380 bp DNA linear BCT 07-AUG-2000  
 LOCUS AF260345 Haemophilus influenzae strain 667 outer membrane protein PI (ompP1)  
 DEFINITION gene, complete cds.  
 ACCESSION AF260345  
 VERSION AF260345.1 GI:9716582  
 KEYWORDS Haemophilus influenzae.  
 SOURCE Haemophilus influenzae.  
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pasteurellales; Haemophilus.

REFERENCE 1 (bases 1 to 1380)  
 AUTHORS Bolduc, G.R., Bouchet, V., Jiang, R.Z., Geisselsoder, J.,  
 Truong-Bolduc, Q.C., Rice, P.A., Peiton, S.I. and Goldstein, R.  
 TITLE Variability of outer membrane protein PI and its evaluation as a  
 vaccine candidate against experimental otitis media due to  
 nontypeable Haemophilus influenzae: an unambiguous, multifaceted  
 approach

JOURNAL Infect. Immun. 68 (8), 4505-4517 (2000)  
 MEDLINE 20359342  
 PUBMED 10899849  
 REFERENCE 2 (bases 1 to 1380)  
 AUTHORS Bouchet, V.  
 TITLE Direct Submission

## JOURNAL

Submitted (25-APR-2000) Section of Molecular Genetics, The Maxwell  
 Finland Laboratory for Infectious Diseases, Boston University  
 School of Medicine, Boston Medical Center, 774 Albany Street,  
 Boston, MA 02118, USA

## FEATURES

## Source

Location/Qualifiers

1..1380

/organism="Haemophilus influenzae"

/strain="667"

/serotype="non-typable"

/db\_xref="taxon:727"

/country="Finland"

1..1380

/gene="ompP1"

1..1380

/gene="ompP1"

/note="OMP p1"

/codon\_start=1

/transl\_table=11

/product="outer membrane protein p1"

/protein\_id="AAF97559.1"

/db\_xref="GI:9716583"

/translation="MKKFNQSLATALLAVGGANAAAFQAEVSTSGLRAYAGEAA

IADNASVATNPALMSLFKTAQFSTGGVYIDSRINMGDVSNDALITSTNSGVKTIKH

GSASQNVIPGAFVFNLYFPVAVNDKALGAGMNVFGLKSEYDDSDYAGIFGKTDL

SAINLNSGATRVTEGUSLGLVNDKAYAKQVERNAGLIVDSLDNQVQTALSVOKEP

LKYLHKYLPKSDTSVSLQDRAAMGFGNAGVYQFNEANRIGLAYHSKVIDIDART

ATSEADYIKAKGKGLDTLTLPDYLESGFQLDKLAVHYSYKTHWSRITKLNASF

EDGKAFDKELQYSNNRISIALGASVLYEKLTLRAGIAYDQAAASHORSAIPDRT

WYSLGATYKFTPNLSVDLGYAYLKGKKVHEKVTIGDERSLALNTTANYTSQAHNL

YGLNLYSF"

BASE COUNT 431 a 235 c 292 g 422 t

## ORIGIN

## Alignment Scores:

Pred. No.: 4,43e-48 Length: 1380  
 Score: 547.00 Matches: 106  
 Percent Similarity: 77.33% Conservative: 27  
 Best Local Similarity: 61.63% Mismatches: 33  
 Query Match: 57.16% Indels: 6  
 DB: 1 Gaps: 3

US-09-995-493-52 (1-179) x AF260345 (1-1380)

QY 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31  
 DB 868 AAGGTGATTAAACCTTACATTCAGCAGATTACTTAGAACCTTTCGGTTCCATTA 927  
 QY 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTyrSerArgPheLysGlu 51  
 DB 928 ACTGACAACTTGGCGTGCATTATAGTATATAATATACCCATTGGAGTCGTTTACAAAA 987  
 QY 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluTyr 71  
 DB 988 TTAACCGTAGCTTCGAAGTGGTAAA-----AAAGCGTTTGATAAGAAATTACAATAC 1041  
 QY 72 LysAspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeu 91  
 DB 1042 AGTAATACTCTCGATTGGCATTAGGGGCAAGTTATATCTTTATGAAAACTCACCTTA 1101  
 QY 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLysLeuSerAlaSerIlePro 111  
 DB 1102 CGTGGGGTATTGCTTAGCATCAACGGCATCTCGTCATCAGCGTAGTGTGCAATTCCA 1161  
 QY 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131  
 DB 1162 GATACCGATCGCACTTGGTATAGTTAGGGGCAACCTATAAATTCACGCCGAATTATCT 1221  
 QY 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151  
 DB 1222 GTTGATCTTGGCTATGCTTACTTAAAGGCAAAAAAGTTTCACTTTAAAGAAATAAACA 1281  
 QY 152 Ile-----LysGlyLeuLeuLeu---ValGluAlaAspTyrThrThrLysAlaThr 167

```

Db 1282 ATAGGTGACGACGTTCACTTCGATTGAATACAACTGCAAAATATATCTTCTCAAGCACAT 1341
QY 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179 linear BCT 07-AUG-2000
||||| 1380 bp DNA outer membrane protein P1 (ompP1)
Db 1342 GCAATCTTTACGGCTTAACTTAATATATAGTTTC 1377

RESULT 37
AF260346 1380 bp DNA linear BCT 07-AUG-2000
LOCUS AF260346 gene, complete cds.
DEFINITION AF260346 Haemophilus influenzae strain 199 outer membrane protein P1 (ompP1)
ACCESSION AF260346
VERSION AF260346.1 GI:9716584
KEYWORDS Haemophilus influenzae.
SOURCE Bacteria: Proteobacteria; gamma subdivision; Pasteurellaceae;
ORGANISM Haemophilus.
REFERENCE 1 (bases 1 to 1380)
AUTHORS Bolduc,G.R., Bouchet,V., Jiang,R.Z., Geisselsoder,J.,
TITLE Truong-Bolduc,Q.C., Rice,P.A., Pelton,S.I. and Goldstein,R.
Variability of outer membrane protein P1 and its evaluation as a
vaccine candidate against experimental otitis media due to
nontypeable Haemophilus influenzae: an unambiguous, multifaceted
approach
JOURNAL Infect. Immun. 68 (8), 4505-4517 (2000)
MEDLINE 20359342
PUBMED 10899849
REFERENCE 2 (bases 1 to 1380)
AUTHORS Bouchet,V.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2000) Section of Molecular Genetics, The Maxwell
Finland Laboratory for Infectious Diseases, Boston University
School of Medicine, Boston Medical Center, 774 Albany Street,
Boston, MA 02118, USA
FEATURES
source
Location/Qualifiers
1..1380
/organism="Haemophilus influenzae"
/strain="199"
/serotype="non-typable"
/db_xref="taxon:727"
/country="Finland"
1..1380
/gene="ompP1"
1..1380
/gene="ompP1"
/feature="OMP P1"
/codon_start=1
/transl_table=11
/product="outer membrane protein P1"
/protein_id="AAF97560.1"
/db_xref="GI:9716585"
/translation="MKKNOSLLATAMLLAAGGANAALAEVSTSGLGRAYAGEAA
IADNASVATPALMSLEKTAQFSTGGVYIDSRINMGDVNSDALITSTNGYKTIKH
GSASORNVPGAFVPLNLFYAPVNDKFAAGAMVNFGLKSEYDDSDYDAGIFGGKTDL
SAINLUNLSGAVFTGSLGLGVNAYAKAQRVAGIIVDSLKDQVOTALSVOKEP
LKYHLKPLSDTSVYSLQDRAAGFGNAGVYQFNEANRIGLKYHSHSLTDLNADRT
ATSEADYKILKAGSGDLTLPLDYLELSGFHOLTDLAVHYSLKYTHWSRLTKLNASF
EDGKAPDKELQYNNSRIALGASVNLVEKLTLRAGIAYDQAAASHORSAALPDRT
WYSLGATYKFTPLNSVDLGLYALKGKKVHFVKVTIGDERSLALNTTANYTSQAHNL
YGLNLNYSF"
BASE COUNT 432 a 237 c 291 g 420 t
ORIGIN
Alignment Scores:
Pred. No.: 4,43e-48 Length: 1380
Score: 547.00 Matches: 106
Percent Similarity: 77.33% Conservative: 27
Best Local Similarity: 61.63% Mismatches: 33
Query Match: 57.16% Indels: 6
DB: 1 Gaps: 3
US-09-995-493-52 (1-179) x AF260346 (1-1380)

```

```

QY 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31
||||| 1371 bp DNA linear BCT 07-AUG-2000
Db 868 AAAGGTGATTTAAACCTTACATTCGACAGATTACTTAAAGAACTTCTCGTTCCATCAATTA 927
||||| 1371 bp DNA linear BCT 07-AUG-2000
QY 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51
||||| 1371 bp DNA linear BCT 07-AUG-2000
Db 928 ACTGACAAACTTCGGCTGCTATAGTTATAATATATACCATTTGGAGTCTGTTTAAACAAA 987
||||| 1371 bp DNA linear BCT 07-AUG-2000
QY 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71
||||| 1371 bp DNA linear BCT 07-AUG-2000
Db 988 TTAACGCTAGCTTCGAAGATGGTAAA-----AAAGCGTTTTCATAAAGAAATTTACAAATAC 1041
||||| 1371 bp DNA linear BCT 07-AUG-2000
QY 72 LysAspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeu 91
||||| 1371 bp DNA linear BCT 07-AUG-2000
Db 1042 AGTAATAACTTCGTATTCATTTAGGGCAAGTTATATCTTTATGAAAACTCACCTTA 1101
||||| 1371 bp DNA linear BCT 07-AUG-2000
QY 92 ArgAlaGlyLeuAlaTyrAspLysAlaIleSerLysThrHisLeuSerAlaSerIlePro 111
||||| 1371 bp DNA linear BCT 07-AUG-2000
Db 1102 CGTGGGGTATTGCTTACGATCAAGCGGCATCTCGTCATCAGCGTAGTCTGCAATTCGA 1161
||||| 1371 bp DNA linear BCT 07-AUG-2000
QY 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131
||||| 1371 bp DNA linear BCT 07-AUG-2000
Db 1162 GATACCGATCGGCACCTTGGTATAGTTAGGGCAACCTATATAATTCACGCCGAATTTATCT 1221
||||| 1371 bp DNA linear BCT 07-AUG-2000
QY 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGlnThrGlnAsn 151
||||| 1371 bp DNA linear BCT 07-AUG-2000
Db 1222 GTTGATCTTGGCTATGCTTACTTAAAGGCAAAAAAGTTTCACCTTTAAAGAACTAAAAACA 1281
||||| 1371 bp DNA linear BCT 07-AUG-2000
QY 152 Ile-----LysGlyLeuLeuLeuLeu-----ValGluAlaAspTyrThrLysAlaThr 167
||||| 1371 bp DNA linear BCT 07-AUG-2000
Db 1282 ATAGGTGACGACGTTCACTTCGATTGAATACAACTGCAAAATATATCTTCTCAAGCACAT 1341
||||| 1371 bp DNA linear BCT 07-AUG-2000
QY 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179 linear BCT 07-AUG-2000
||||| 1371 bp DNA linear BCT 07-AUG-2000
Db 1342 GCAATCTTTACGGCTTAACTTAAATATATAGTTTC 1377
||||| 1371 bp DNA linear BCT 07-AUG-2000
RESULT 38
AF260359 1371 bp DNA linear BCT 07-AUG-2000
LOCUS AF260359 Haemophilus influenzae strain BCH-2 outer membrane protein P1
DEFINITION AF260359 (ompP1) gene, complete cds.
ACCESSION AF260359
VERSION AF260359.1 GI:9716610
KEYWORDS Haemophilus influenzae.
SOURCE Bacteria: Proteobacteria; gamma subdivision; Pasteurellaceae;
ORGANISM Haemophilus.
REFERENCE 1 (bases 1 to 1371)
AUTHORS Bolduc,G.R., Bouchet,V., Jiang,R.Z., Geisselsoder,J.,
TITLE Truong-Bolduc,Q.C., Rice,P.A., Pelton,S.I. and Goldstein,R.
Variability of outer membrane protein P1 and its evaluation as a
vaccine candidate against experimental otitis media due to
nontypeable Haemophilus influenzae: an unambiguous, multifaceted
approach
JOURNAL Infect. Immun. 68 (8), 4505-4517 (2000)
MEDLINE 20359342
PUBMED 10899849
REFERENCE 2 (bases 1 to 1371)
AUTHORS Bouchet,V.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2000) Section of Molecular Genetics, The Maxwell
Finland Laboratory for Infectious Diseases, Boston University
School of Medicine, Boston Medical Center, 774 Albany Street,
Boston, MA 02118, USA
FEATURES
source
Location/Qualifiers
1..1371
/organism="Haemophilus influenzae"
/strain="BCH-2"
/serotype="non-typable"
/db_xref="taxon:727"
/country="USA"

```



```

Db 922 GACAACTGCGACTGATATAGTTTAAATATATACCATTTGGAGTCGTTTACAAAATTA 981
QY 53 ArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyrLys 72
Db 982 AACGCTAGCTTCGAAGATGGTAAG-----AAAGCGTTTGATAGAATTAACAATACAGT 1035
QY 73 AspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAlaLeuThrLeuArg 92
Db 1036 AATAACTCTCGTTCATAGGGCAAGTTATACCTTGAACAAATGACCTTACCT 1095
QY 93 AlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIleProAsp 112
Db 1096 GCGGTATTGCTTACCATCAAGCGGATCTCGTCATCAGCGTAGCTGCAATTCAGAT 1155
QY 113 ThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSerVal 132
Db 1156 ACCGATCGCACCTGGTATAGTTTAGGGCAACCTATAAATTCACCCCTAATTTATCTGTT 1215
QY 133 AspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsnIle 152
Db 1216 GATCTTGGCTAGCTTACTTAAAGGCAAAAAGTTTCACTTTAAAGAAAGTAAACAAATA 1275
QY 153 -----LysGlyLeuLeuLeu-----ValGluAlaAspTyrThrThrLysAlaThrAla 168
Db 1276 GGTGCAAAACGTTTCACTTGCATTTAGTACAACTTATACCTTCTCAAGCACACGCA 1335
QY 169 AsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179
Db 1336 AATCTTTACGGTTTGAATTTAAATATATAGTTTC 1368

RESULT 40
AF260339 1371 bp DNA linear BCT 07-AUG-2000
LOCUS Haemophilus influenzae strain 3-H-1552 outer membrane protein p1
DEFINITION (omp1) gene, complete cds.
ACCESSION AF260339
VERSION AF260339.1 GI:9716570
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
Vaccine candidate against experimental otitis media due to nontypeable Haemophilus influenzae: an unambiguous, multifaceted approach
JOURNAL Infect. Immun. 68 (8), 4505-4517 (2000)
MEDLINE 20359342
PUBMED 10899849
REFERENCE 2 (bases 1 to 1371)
AUTHORS Bouchet, V.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2000) Section of Molecular Genetics, The Maxwell Finland Laboratory for Infectious Diseases, Boston University School of Medicine, Boston Medical Center, 774 Albany Street, Boston, MA 02118, USA
FEATURES
source
1..1371
/organism="Haemophilus influenzae"
/strain="3-H-1552"
/serotype="non-typable"
/db_xref="taxon:727"
/country="USA"
1..1371
/gene="omp1"
1..1371
/gene="omp1"
/feature="OMP p1"
/codon_start=1
/transl_table=11

```

```

/product="outer membrane protein p1"
/protein_id="AAF97553.1"
/db_xref="GI:9716571"
/translation="MKKEFNSQSLATAMLLAAGGANAFAAEVTSGLGRAYAGEAA
IADNASVATNPALMSLFTNQFSGVGVYVDSRINMGDVSYALISDSIKYTNDSGA
SARNVPGAFVPLNYFVAPNDKFAAGMNVNFKLSEYDDSDYAGVFGGKTDULTAI
NLNLSGAYRTVEGLSLGAVNAQVAKQVNRNAGLIADTVKDNQVKNVLTVOEPLKY
INKYLPKSDTSVSLQDRAAGVFGWAGVYQFNEANRIGLAYHSKVDLTFDRTATS
LEANYINASKQYDGLTLTLPDYLELSGHLQDKLAVHYSYKYTHSRILTCLNASPEDG
KKAPDKELQYNNNSRVALLGASYNLDEKLTLAGIAYDOAASRHOESAAIPDRTWYS
LCATYKFTPNLSVDLGYALYLGKKVHFKEVTIGDKRSLALSTTANYTSSQAHNLGL
NLNVSF"
BASE COUNT 424 a 233 c 289 g 425 t
ORIGIN
Alignment Scores:
Pred. No.: 9,08e-48 Length: 1371
Score: 544.00 Matches: 105
Percent Similarity: 77.78% Conservatives: 28
Best Local Similarity: 61.40% Mismatches: 32
Query Match: 56.84% Indels: 6
DB: 1 Gaps: 3
US-09-995-493-52 (1-179) x AF260339 (1-1371)
QY 13 GlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThr 32
Db 862 GGTGATTTAACCTTACATTCGCAGATTACTTAGAATCTTCTGGTTCCATCAATTA 921
QY 33 AspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGluLeu 52
Db 922 GACAACTTGCAGTCGATATAGTTTAAATATATACCATTTGGAGTCGTTTACAAAATTA 981
QY 53 ArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyrLys 72
Db 982 AACGCTAGCTTTCGAAGATGGTAAG-----AAAGCGTTTGATAGAATTAACAATACAGT 1035
QY 73 AspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAlaLeuThrLeuArg 92
Db 1036 AATAACTCTCGTTCATAGGGCAAGTTATACCTTGAACAAATGACCTTACCT 1095
QY 93 AlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIleProAsp 112
Db 1096 GCGGTATTGCTTACCATCAAGCGGATCTCGTCATCAGCGTAGCTGCAATTCAGAT 1155
QY 113 ThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSerVal 132
Db 1156 ACCGATCGCACCTGGTATAGTTTAGGGCAACCTATAAATTCACCCCTAATTTATCTGTT 1215
QY 133 AspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsnIle 152
Db 1216 GATCTTGGCTAGCTTACTTAAAGGCAAAAAGTTTCACTTTAAAGAAAGTAAACAAATA 1275
QY 153 -----LysGlyLeuLeuLeu-----ValGluAlaAspTyrThrThrLysAlaThrAla 168
Db 1276 GGTGCAAAACGTTTCACTTGCATTTAGTACAACTTATACCTTCTCAAGCACACGCA 1335
QY 169 AsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179
Db 1336 AATCTTTACGGTTTGAATTTAAATATATAGTTTC 1368
Search completed: May 19, 2003, 19:39:11
Job time : 1747 secs

```



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using 'frame\_plus.p2n.model'

Run on: May 19, 2003, 18:28:18 ; Search time 172 Seconds  
(without alignments)  
2343.649 Million cell updates/sec

Title: US-09-995-493-52

Perfect score: 957

Sequence: 1 QHNGVGPYIGKSLTLKP.....ADYTKATANLYGLNLNRYF 179

Scoring table:

BLOSUM62 Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delopt 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool-QFMR59493/runat\_12052003.091308.22995/app\_query.fasta\_1.327  
-DB=N\_Geneseq\_101002 -QFMR=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0995493 @cgn\_1.1.200 @runat\_12052003.091308.22995 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_101002.\*

1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	549	57.4	1598	17	AAT28520	H. influenzae dete
2	549	57.4	1598	22	ABA76850	Haemophilus influe
3	549	57.4	1830121	17	AAT42063	Haemophilus influe
4	256	26.8	240	12	AAQ12083	H. influenzae strai
5	151	15.8	1383	21	AAZ54519	Neisseria gonorrhoe
6	144.5	15.1	1329	21	AAA49619	Neisseria meningit
7	144.5	15.1	1401	21	AAZ54520	Neisseria meningit
8	144.5	15.1	65632	21	AAAB1502	N. meningitidis pa
C	144.5	15.1	349980	21	AAF21544	Neisseria meningit
	143.5	15.0	1401	21	AAA49618	Neisseria meningit
	142.5	14.9	1395	21	AAZ54521	Neisseria meningit
	133.5	13.9	1371	24	ABA97241	tmoX nucleotide se
	133.5	13.9	1362	24	ABA97235	tmoX gene nucleoti
	140	11.3	1764	22	AAF25600	H. pylori HP5188 e
	102	10.7	1068	19	AAK30610	H. pylori outer me
	102	10.7	1764	19	AAK30611	H. pylori outer me
	98	10.2	1941	22	AAK3872	Neisseria meningit
	98	10.2	1941	22	AAK17037	N. meningitidis st
18	96.5	10.1	2127	21	AAZ54328	Neisseria meningit
20	96.5	10.1	56485	21	AAAB1476	N. meningitidis pa
C	96.5	10.1	349980	21	AAF21612	Neisseria meningit
C	96.5	10.1	837096	21	AAAB1489	N. meningitidis pa
23	95.5	10.0	2169	21	AAK59216	DNA encoding a Nel
24	95	9.9	1746	22	AAK3869	Neisseria meningit
25	95	9.9	1746	22	AAK17034	N. meningitidis st
26	94.5	9.9	1719	24	ABQ68105	Listeria monocytog
27	94.5	9.9	1719	24	ABQ69946	Listeria monocytog
28	93.5	9.8	2078	21	AAK59217	DNA encoding a Nel
29	93	9.7	1035	24	ABA91422	Haemophilus paraga
30	93	9.7	1035	24	ABA91425	Haemophilus paraga
31	92.5	9.7	465	21	AAZ54327	Neisseria gonorrhoe
32	90	9.4	1035	24	ABA91417	Haemophilus paraga
33	90	9.4	1035	24	ABA91418	Haemophilus paraga
34	90	9.4	1035	24	ABA91420	Haemophilus paraga
35	89.5	9.4	1026	24	ABA91421	Haemophilus paraga
36	89.5	9.4	1026	24	ABA91424	Haemophilus paraga
37	89.5	9.4	1026	24	ABA91427	Haemophilus paraga
38	89	9.3	1650	17	AAK04376	Moraxella catarrha
39	89	9.3	1650	20	AAZ09296	M. catarrhalis out
40	89	9.3	119211	22	AAF28553	Genomic fragment #
41	88.5	9.2	28730	21	AAA81486	N. meningitidis pa
42	88	9.2	1035	24	ABA91419	Haemophilus paraga
43	88	9.2	1035	24	ABA91423	Haemophilus paraga
C	88	9.2	20757	20	AAK20599	Polynucleotide seq
	87	9.1	1035	24	ABA91426	Haemophilus paraga

ALIGNMENTS

RESULT 1

AAT28520

ID AAT28520 standard; DNA; 1598 BP.

XX

AC AAT28520;

XX

DT 02-APR-1997 (first entry)

XX

DE H. Influenzae detection probe #2.

XX

Detection; probe; amplification primer; bacterial pathogen; pneumonia;  
Escherichia coli; Klebsiella pneumoniae; pseudomonas aeruginosa;  
Kw Proteus mirabilis; Streptococcus pneumoniae; Staphylococcus aureus;  
Kw Staphylococcus epidermidis; Enterococcus faecalis; respiratory tract;  
Kw Staphylococcus saprophyticus; Streptococcus pyogenes; urinary tract;  
Kw Haemophilus influenzae; Moraxella catarrhalis; septicaemia; meningitis;  
Kw infection; intra-abdominal infection; skin infection;  
Kw bacterial resistance; beta-lactam antibiotic; ds.

XX OS Synthetic.  
 XX PN WO9608582-A2.  
 XX PD 21-MAR-1996.  
 XX PF 12-SEP-1995; 95WO-CA00528.  
 XX PR 12-SEP-1994; 94US-0304732.  
 XX PA (BERG/) BERGERON M G.  
 XX PA (QUEL/) QUELLETTE M.  
 XX PA (ROYP/) ROY P H.  
 XX PI Bergeron MG, Ouellette M, Roy PH;  
 XX DR WPI; 1996-179953/18.  
 XX PT Method for the detection of bacterial species using probes and  
 PT primers - allows detection and quantification of antibiotic  
 PT resistant bacteria in patients, the environment and food  
 XX PS Claim 47; Page 80-81; 216pp; English.  
 XX CC The sequences given in AAT28519-24 represent probes which were used in  
 CC the method of the invention for the detection of H. influenzae in a  
 CC sample. The method comprises using probes and/or amplification primers  
 CC which are specific, ubiquitous and sensitive for determining the  
 CC presence and/or amount of nucleic acids from selected bacterial species  
 CC in any sample, where the bacterial nucleic acid comprises a selected  
 CC target region hybridisable with the probes or primers. The method  
 CC comprises contacting the sample with the probes or primers and detecting  
 CC the presence and/or amount of hybridised primers or amplification  
 CC products as and indication of the presence and/or amount of the  
 CC bacterial species. This method may be used to detect commonly  
 CC encountered bacterial pathogens, e.g. Escherichia coli, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa, Proteus mirabilis, Streptococcus  
 CC pneumoniae, Staphylococcus aureus, Staphylococcus epidermidis,  
 CC Enterococcus faecalis, Staphylococcus saprophyticus, Streptococcus  
 CC pyogenes, Haemophilus influenzae and Moraxella catarrhalis. These  
 CC bacterial species are associated with approx. 90% of urinary tract  
 CC infections and with a high percentage of other severe infections,  
 CC including septicaemia, meningitis, pneumonia, intra-abdominal infections,  
 CC skin infections and other severe respiratory tract infections. The  
 CC method may also be used to evaluate a bacterial resistance to  
 CC beta-lactam antibiotics.  
 XX SQ Sequence 1598 BP; 500 A; 278 C; 329 G; 491 T; 0 other;

Alignment Scores:  
 Pred. No.: 6.67e-59 Length: 1598  
 Score: 549.00 Matches: 105  
 Percent Similarity: 76.74% Conservative: 27  
 Best Local Similarity: 61.05% Mismatches: 34  
 Query Match: 57.37% Indels: 6  
 DB: 17 Gaps: 2

US-09-995-493-52 (1-179) x AAT28520 (1-1598)

QY 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31  
 DB 939 AAAGGTAAATTAACTTTACATGCCAGATTACATTAAGCACTTCTGGTTTCCATCAATTA 998  
 QY 32 ThrAspGlnTrpAlaLeuHisTyrSerTyrTyrThrGluTrpSerArgPheLysGlu 51  
 DB 999 ACTGACAACTTCGAGTGCATATAGTTATATAATATATACCCATGGAGTCGTTTAAACAAA 1058  
 QY 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71  
 DB 1059 TTATCATCCAGCTCGAAGATGGTAA- - - - -AAAGCTTTTGAAGAAGATTACAATAC 1112  
 QY 72 LysAspAsnSerArgPheAlaLeuGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeu 91

DB 1113 AGTAATAACTCTCGTGTTCATAGGGGCAAGTTATATCTTTATGAAAAATTCACCTTA 1172  
 QY 92 ArgAlaGlyLeuAlaTyrAspLysAlaLeuSerLysThrHisLeuSerAlaSerIlePro 111  
 DB 1173 CGTGGGGGTATTGCTTACGATCAAGCGGCATCTCGTCATCACCGTAGCTGCAATTC 1332  
 QY 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131  
 DB 1233 GATACCGATCGCACTTGGTATAGTTAGTGCAACCTATAAATTCACGCCGAATTTATCT 1292  
 QY 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151  
 DB 1293 GTTCATCTGGCTATGCTTACTTTAAAGGCAAAAGTTCACCTTTAAAGGAAGTAAAAACA 1352  
 QY 152 IleLysGly- - - - -LeuLeuLeuValGluAlaAspTyrThrLysAlaThr 167  
 DB 1353 ATAGTGACAAACGTACATTCACATTCGAATCAACTGCAATATATATCTTCTCAAGCACAC 1412  
 QY 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
 DB 1413 GCAAACTCTTACGGTTTGAATTTAAATATAGTTTC 1448

RESULT 2  
 ABA76850  
 ID ABA76850 standard; DNA; 1598 BP.  
 XX AC ABA76850;  
 XX DT 28-JAN-2002 (first entry)  
 XX DE Haemophilus influenzae polynucleotide SEQ ID NO 26.  
 XX KW Detection: bacterial species; animal; food; environment;  
 XX KW antibiotic resistance; ds.  
 XX OS Haemophilus influenzae.  
 XX PN NZ501596-A.  
 XX PD 29-JUN-2001.  
 XX PF 12-SEP-1995; 95NZ-0501596.  
 XX PR 12-SEP-1995; 95NZ-0501596.  
 XX PA (IDII-) IDI INFECTIO DIAGNOSTIC INC.  
 XX PI Bergeron MG, Ouellette M, Roy PH;  
 XX DR WPI; 2001-615034/71.  
 XX PT Method for detecting target bacterial species in a sample, comprises  
 PT detecting the presence or amount of bacterial nucleic acid amplified by  
 PT a primer derived from bacterial DNA, specific for the target bacterial  
 PT species  
 XX PS Claim 6; Page 95-96; 168pp; English.  
 XX CC The invention relates to detecting target bacterial species suspected to  
 CC be present in a sample, comprising contacting nucleic acids of target  
 CC bacterial species with an amplification primer pair derived from a  
 CC bacterial DNA fragment (ABA76825-ABA76861) specific for the target  
 CC bacterial species but ubiquitous for different strains, amplifying the  
 CC nucleic acid and detecting the presence or amount of an amplified  
 CC sequence as an indication of the presence or amount of the target  
 CC bacterial species. The invention includes primers and probes  
 CC (ABA76862-ABA76984) against the target bacterial species, especially  
 CC E.coli, K.pneumoniae, P.aeruginosa, P.mirabilis, S.pneumoniae,  
 CC S.aureus, S.epidermidis, E.faecalis, S.saprophyticus, S.pyogenes,  
 CC H.influenzae, M.catarrhalis and/or group A Streptococci producing  
 CC exotoxin A gene A, suspected to be present in a sample which is  
 CC obtained from human patients, animals, environment or food, and which



CC consists of one or more bacterial colonies. Oligonucleotide  
 CC probes and primers complementary to the bacterial genes encoding  
 CC resistance to antibiotics such as bla(tem), bla(rob), bla(shv), aadB,  
 CC aacC1, aacC2, aacC3, aacC4, mecA, vanA, vanH, vanX, sacA, aacA-aphD, vat,  
 CC vga, msrA, sul and/or int (ABA76985-ABA77001) are also useful to identify  
 CC commonly encountered and clinically important resistance genes. The  
 CC invention provides a rapid method of bacterial identification that can be  
 CC achieved, which reduces the time currently required for the  
 CC identification of pathogens in the clinical laboratory.

XX Sequence 1598 BP; 500 A; 278 C; 329 G; 491 T; 0 other;

## Alignment Scores:

Pred. No.: 6.67e-59 Length: 1598  
 Score: 549.00 Matches: 105  
 Percent Similarity: 76.74% Conservatives: 27  
 Best Local Similarity: 61.05% Mismatches: 34  
 Query Match: 57.37% Indels: 6  
 DB: 22 Gaps: 2

US-09-995-493-52 (1-179) x ABA76850 (1-1598)

QY 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31  
 Db 939 AAGGTAATTTACCTTACCTTACCTGCCAGATTACTTAGAATTTCTGGTTTCCATCAATTA 998  
 QY 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51  
 Db 999 ACTGACAAACTTCGAGTCGATATAGTTATATAATATACCATTTGGAGTCGTTTAAACANAA 1058  
 QY 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71  
 Db 1059 TTACATGCCAGCTTCGAAGATGGTAA-----AAAGCTTTTGTAAAGAAATTACATAC 1112  
 QY 72 LysAspAsnSerArgPheAlaIleGlyThrTyrTyrSerLeuAsnAspAlaLeuThrLeu 91  
 Db 1113 AGTAATAACTCTCGTGTGCTAGGAGGCAAGTTATATATCTTTATGAAAATTGACCTTA 1172  
 QY 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111  
 Db 1173 CGTGGGGTATGCTTACGATCAAGCGCATCTCGTCATCACCGCTAGTGTGCAATTCCTCA 1232  
 QY 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131  
 Db 1233 GATACCGATCGCACTTGGTATAGTTAGGTGCAACCTTATAAATTCAGCGCAATTTATCT 1292  
 QY 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151  
 Db 1293 GTTGATCTTGGCTATGCTTACTTAAAGGCAAAAAGTTCCATTAAGAAGTAAACACA 1352  
 QY 152 IleLysGly-----LeuLeuValGluAlaAspTyrThrThrLysAlaThr 167  
 Db 1353 ATAGGTGACAAAGGTACATTTGACATTTGAATACAACTGCAAAATTATCTTCTCAAGCAC 1412  
 QY 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
 Db 1413 GCAATCTTACGGTTTGAATTTAAATATATAGTTTC 1448

RESULT 3

AAT42063

ID AAT42063 standard; DNA; 1830121 BP.

XX

AC AAT42063;

XX

DT 14-SEP-1999 (first entry)

XX

DE Haemophilus influenzae complete genome sequence.

XX

KW Genome; bacterium; Haemophilus influenzae; computer readable medium;  
 expression modulating fragment; regulation; gene expression; vector;  
 organism; open reading frame; ORF; ds.

XX

OS Haemophilus influenzae.

XX WO9633276-A1.

XX 24-OCT-1996.

XX

PD

XX

PF 22-APR-1996; 96WO-US05320.

XX

XX 07-JUN-1995; 95US-0487429.

XX

PR 21-APR-1995; 95US-0426787.

XX

PR 07-JUN-1995; 95US-0476102.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

PA (UYJO ) UNIV JOHNS HOPKINS.

XX

XX Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;

XX

XX WPI; 1996-485782/48.

XX

XX Haemophilus influenzae Rd genome recorded on computer readable

PT medium - useful for identifying commercially important nucleic acid

PT fragments by homology searching

XX

XX Claim 1; Page 77.2-77.1091; 1291pp; English.

XX

XX This sequence represents the complete genome sequence of the bacterium

CC Haemophilus influenzae strain Rd. The invention relates to a computer

CC readable medium (CRM) having recorded upon it the complete H.influenzae

CC nucleotide sequence (1), a representative fragment of (1) or a nucleotide

CC sequence at least 99% identical to (1). By providing the full-length

CC genomic sequence in a computer readable form, it is possible to identify

CC commercially important nucleic acid fragments and expression modulating

CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to

CC regulate the expression of a nucleic acid molecule. Vectors and altered

CC organisms comprising the predicted ORFs can be used to produce any of the

CC polypeptide fragments of the H. influenzae Rd genome.

XX

XX Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

XX

XX Alignment Scores:

Pred. No.: 1.26e-54 Length: 1830121

Score: 549.00 Matches: 106

Percent Similarity: 77.91% Conservatives: 28

Best Local Similarity: 61.63% Mismatches: 32

Query Match: 57.37% Indels: 6

DB: 17 Gaps: 3

US-09-995-493-52 (1-179) x AAT42063 (1-1830121)

QY 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31

Db 422414 AAAGGTGATTAAACCCCTTACATTTGCCAGATTACTTAGAATTTCTGGTTTCCATCAATTA 422473

QY 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51

Db 422474 ACTGACAAACTTGGTGTGCAATTTATAGTTATATAATATACCATTTGGAGTCGTTTAAACANAA 422533

QY 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71

Db 422534 TTAACGCTAGCTCGAAGATGGTAAA-----AAAGCTTTTGTAAAGAATTACAAATAC 422587

QY 72 LysAspAsnSerArgPheAlaIleGlyThrTyrTyrSerLeuAsnAspAlaLeuThrLeu 91

Db 422588 AGTAATAACTCTCGTGTGCTAGGAGGCAAGTTATATCTTTGATGAAAATTGACCTTA 422647

QY 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111

Db 422648 CGTGGGGTATGCTTACGATCAGGCTGCATCTCGTCATCAGCGTAGTGTGCAATTCCTCA 422707

QY 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131

Db 422708 GATACCGATCGCACTTGGTATAGTTAGGTGCAACCTTATAAATTCACCGCAATTTATCT 422767

QY 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151



CC have use as antibacterial agents. The polynucleotides of the invention  
XX may also be used in gene therapy protocols.

SQ Sequence 1383 BP; 379 A; 450 C; 314 G; 240 T; 0 other;

## Alignment Scores:

Pred. No.: 6.21e-09 Length: 1383  
Score: 151.00 Matches: 43  
Percent Similarity: 40.93% Conservativeness: 36  
Best Local Similarity: 22.28% Mismatches: 74  
Query Match: 15.78% Indels: 40  
DB: 21 Gaps: 5

US-09-995-493-52 (1-179) x AA254519 (1-1383)

QY 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31  
DB 844 AAAGCCAGTCTCAAAATCGTACCCCTGAGTCTTTGCGTACACGGCATGTACAAAGTG 903  
QY 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLys--- 50  
DB 904 TCGCACAAAGCCGACCTGTTCCGGACGCTAACTTGGACGCCACACCGCTTCAATAAG 963  
QY 51 -----GluLeuArgGlyLysTyrClnAspGlySerGlyTyrGluAlaPhe 65  
DB 964 CGGGAAGTGTCTTTTGAAGAAAGAAATAATATTCTAATGCGCAAAATCCGACCGCAC 1023  
QY 66 ThrLysLysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThrTyrSerLeu 85  
DB 1024 ACCATCACCCCACTGCGGCAACACCTACAAAGTCGCTGGCGTCTTATCAATC 1083  
QY 86 AsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLys-----AlaAlaSerLys 103  
DB 1084 AGGCAACCGTGCACCTGCGGCTCGCATCGCTTTTGACAAACCGCTGTCCGCAACGCC 1143  
QY 104 ThrHisLeuSerAlaSerIleProAspThrAspArgMetTrpTyrSerIleGlyAlaThr 123  
DB 1144 GACTACCGCATGACAGCTGCGCGGCAACCGCATCTGTTCTCCGCGGCATGAAA 1203  
QY 124 TyrLysPheThrProAsnLeuSerValAspValGlyPheAlaHisLeuArgGlyLysLys 143  
DB 1204 TACCATATCGGCAAAACACCGCTGCTGATCGCCCTACACCCACATC-----1251  
QY 144 LysHisPheValGluThrGlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThr 163  
DB 1252 ---CACATCAACGACAC-----AGCTACCGCACG 1278  
QY 164 ThrLysAlaThrAlaAsn-----169  
DB 1279 CGGAAGGCAACGCGCAACGATGTGGACAGCAAGGTCGCTTTCGCGCACGTTTCAAAAC 1338  
QY 170 -----LeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
DB 1339 CACCGCCACATCATCGCGCTGCATACATACACCTACAAATTC 1377

## RESULT 6

AAA49619  
ID AAA49619 standard; DNA; 1329 BP.

AC AAA49619;

DT 27-OCT-2000 (first entry)

DE Neisseria meningitidis BASB044 gene #2.

KW Meningitis; microbial disease; upper respiratory tract infection;  
KW bacteraemia; invasive bacterial disease; BASB044 gene; ds.

OS Neisseria meningitidis.

PH Key Location/Qualifiers

FT CDS 1..1329

FT /\*tag= a

FT /product= "BASB044 protein"  
XX /partial

PN WO200034482-A2.

XX 15-JUN-2000.

XX 07-DEC-1999; 99WO-IB02014.

XX 08-DEC-1998; 98GB-0026979.

XX 08-DEC-1998; 98GB-0026980.

XX 17-DEC-1998; 98GB-0028015.

XX 05-JAN-1999; 99GB-0000090.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX Ruelle J, Verlant VGCL;

XX WPI; 2000-423427/36.

XX P-PSDB; AAB01259.

XX Novel BASB041, 43, 44 and 48 polypeptides of Neisseria meningitidis

XX useful for diagnostic, prophylactic and therapeutic purposes against

XX microbial diseases comprise a specific amino acid sequence

XX Claim 59; Page 165-166; 171pp; English.

XX The present sequence is the BASB044 gene from the Neisseria

XX meningitidis strain H44/76. The gene and its protein can be used in

XX the prevention and treatment of microbial disease such as bacteraemia,

XX meningitis and upper respiratory tract infections. They are particularly

XX useful for treating bacterial diseases. They can also be used for

XX diagnosing these diseases. The sequence shows strong homology to

XX the BASB044 gene from N. meningitidis strain ATCC 13090.

XX SQ Sequence 1329 BP; 357 A; 427 C; 323 G; 222 T; 0 other;

XX Alignment Scores:

Pred. No.: 3.86e-08 Length: 1329

Score: 144.50 Matches: 45

Percent Similarity: 43.23% Conservativeness: 38

Best Local Similarity: 23.44% Mismatches: 70

Query Match: 15.10% Indels: 39

DB: 21 Gaps: 7

US-09-995-493-52 (1-179) x AAA49619 (1-1329)

QY 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31

DB 793 AAAGCCCGCTTAAATCGTTACGCTTGTTCGCTACACGCTATGTACAAAGTG 852

QY 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPhe----- 49

DB 853 TCCGATAAAGCCGACCTGTTCCGGCAGCTAACTTGGAGCGCCACACCGCTTCGATAAG 912

QY 50 -----LysGlu-----LeuArgGlyLysTyrGlnAspGlySer 60

DB 913 GCGGAAGTGTCTTTTGAAGAAAGAAACCGCTCTCAAGGCAAA-----957

QY 61 GlyTyrGluAlaPheThrLysLysGluGluTyrLysAspAsnSerArgPheAlaIleGly 80

DB 958 ---TCCGACCGCACACCATCACCCTGCGGCAACACCTACAAAGTCCGCTTCGGC 1014

QY 81 ThrThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAla 100

DB 1015 GGTCTTATCAATCAGCAACCGCTGCAACTGCGCGCGCATCGCTTTGACAAATCG 1074

QY 101 AlaSerLysThr-----HisLeuSerAlaSerIleProAspThrAspArgMetTrpTyr 118

DB 1075 CCCGTCGCAACCGCGCTACCGCATACAGAGCTACCCGACCGCACCGCATCTGGTTC 1134

QY 119 SerIleGlyAlaThrTyrLysPheThrProAsnLeuSerValAspValGlyPheAlaHis 138

Db 1135 TCCGCGGTATGAATACCATATTCGGTAAACACAGCTCGTCGATGCCCGCTACACCCAC 1194  
 Qy 139 LeuArgGlyLysLysLysHisPheValGluThrGln----- 150  
 Db 1195 ATC-----CACATCAACGACACACAGCTACCGCACGGGGAAGCAAGCGGC 1239  
 Qy 151 -----AsnIleLysGlyLeuLeuValGluAlaAspTyrThrThrLysAlaThr 167  
 Db 1240 AACGATGTGGACAGCAAGGC-----GGCTCTTCGCGACGTTTCAAAAACACAC 1287  
 Qy 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
 Db 1288 GCGGACATCATCGTCTGCAATACACCTACAAATTC 1323  
 RESULT 7  
 ID AA254520  
 AA254520 standard; DNA; 1401 BP.  
 AC AA254520;  
 XX  
 DT 21-MAR-2000 (first entry)  
 DE Neisseria meningitidis ORF 989 partial DNA; sequence SEQ ID NO:2987.  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;  
 KW antibacterial; gene therapy; ds.  
 XX  
 OS Neisseria meningitidis.  
 PN WO9957280-A2.  
 XX  
 PD 11-NOV-1999.  
 PF 30-APR-1999; 99WO-US09346.  
 PR 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 09-OCT-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;  
 PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX  
 DR WPI; 2000-062150/05.  
 DR P-PSDB; AAY7578.  
 XX  
 PT Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics  
 XX  
 PS Claim 7; Page 1404; 1453pp; English.  
 XX  
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941  
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.

XX  
 SQ Sequence 1401 BP; 371 A; 457 C; 337 G; 236 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 4.15e-08 Length: 1401  
 Score: 144.50 Matches: 45  
 Percent Similarity: 43.23% Conservative: 38  
 Best Local Similarity: 23.44% Mismatches: 70  
 Query Match: 15.10% Indels: 39  
 DB: 21 Gaps: 7  
 US-09-995-493-52 (1-179) x AA254520 (1-1401)  
 Qy 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31  
 Db 865 AAAGCCCGCGTTAAATCGTTACGCTGTAGTCTTTGTCGTACACGGTATGTACAAGTG 924  
 Qy 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPhe----- 49  
 Db 925 TCCGATAAAGCGGACCTGTTCGGCGACGTAACCTTGACGCCGACAGCGCTTCGATAAG 984  
 Qy 50 -----LysGlu-----LeuArgGlyLysTyrGlnAspGlySer 60  
 Db 985 GCGGAACCTGGTTTGAAGAAAGAAAAACCGCTCGTCAAGGCAAA----- 1029  
 Qy 61 GlyTyrGluAlaPheThrLysLysGluLysTyrLysAspAsnSerArgPheAlaIleGly 80  
 Db 1030 ---TCCGACCGCACCCATCACCCCAACTGCCGCAACACCTACAAAGTCGGCTTCGGC 1086  
 Qy 81 ThrThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAla 100  
 Db 1087 GGTCTTATCAANTCAGCAACCGCTGCAACTGCCGCGCGCATCGCTTTTGACAAATCG 1146  
 Qy 101 AlaSerLysThr-----HisLeuSerAlaSerIleProAspThrAspArgMetTrpTyr 118  
 Db 1147 CCCGTCCGCAACGCGGACTACCGCATGAACAGCTTACCAGCGGCAACCGCATCTGGTTC 1206  
 Qy 119 SerIleGlyAlaThrTyrLysPheThrProAsnLeuSerValAspValGlyPheAlaHis 138  
 Db 1207 TCCGCGGTATGAATAACCATATCGTATGTAACCAACACCGTCTGTCGATCGCGCTTACACCCAC 1266  
 Qy 139 LeuArgGlyLysLysLysHisPheValGluThrGln----- 150  
 Db 1267 ATC-----CACATCAACGACACACCACTACCGCACGGGGAAGCAAGCGGC 1311  
 Qy 151 -----AsnIleLysGlyLeuLeuValGluAlaAspTyrThrThrLysAlaThr 167  
 Db 1312 AACGATGTGGACAGCAAGGC-----GGCTCTTCGCGACGTTTCAAAAACACAC 1359  
 Qy 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
 Db 1360 GCGGACATCATCGGCTCGAATACACCTACAAATTC 1395  
 RESULT 8  
 AAA81502  
 ID AAA81502 standard; DNA; 65632 BP.  
 XX  
 AC AAA81502;  
 XX  
 DT 04-DEC-2000 (first entry)  
 DE N. meningitidis partial DNA sequence gnm\_49 SEQ ID NO:49.  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB; ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO200022430-A2.  
 XX  
 PD 20-APR-2000.  
 XX



CC AAF21606 represent PCR primers which are used in the exemplification of  
 CC the present invention. The NMB genome and fragments from it have  
 CC antibacterial activity, and can be used in vaccines and gene therapy.  
 CC Neisseria nucleic acids, proteins and/or antibodies which binds to the  
 CC proteins can be used in compositions for treating or preventing infection  
 CC due to Neisseria bacteria or as a diagnostic reagent for detecting the  
 CC presence of Neisseria bacteria or of antibodies raised to Neisseria  
 CC bacteria. Computers, computer memory, computer storage medium or computer  
 CC databases can be used in a search to identify open reading frames (ORFs)  
 CC or coding sequences within the NMB genome. The DNA sequences provide  
 CC further opportunities to find antigenic or immunogenic proteins which are  
 CC more effective in vaccines than the outer membrane proteins currently  
 CC used.

XX SQ Sequence 349980 BP; 83241 A; 85091 C; 95206 G; 86442 T; 0 other;

Alignment Scores:  
 Pred. No.: 9.32e-05 Length: 349980  
 Score: 144.50 Matches: 45  
 Percent Similarity: 43.23% Conservative: 38  
 Best Local Similarity: 23.44% Mismatches: 70  
 Query Match: 15.10% Indels: 39  
 DB: 21 Gaps: 7

US-09-995-493-52 (1-179) x AAF21544 (1-349980)

Qy 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31  
 Db 100306 AAGCCGCGTTAAATCGTTACGCTGAGTCTTTGTCGGTACACGCTATGTACAAAGTG 100247  
 Qy 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPhe----- 49  
 Db 100246 TCCGATAAAGCCGACCTGTCGGCGACGTAACCTTGGACGCCGCCACGCGCTTCGATAAG 100187  
 Qy 50 -----LysGlu-----LeuArgGlyLysTyrGlnAspGlySer 60  
 Db 100186 CGGCACTGGTTTGAAGAAGAAAAACCGCTCGTCAAGGCCAA----- 100142  
 Qy 61 GlyTyrGluAlaPheThrLysLysGluTyrLysAspAsnSerArgPheAlaIleGly 80  
 Db 100141 ---TCCGACCCGACCCATCACCCCACTGGCGGCAACACCTACAAAGTCGGCTTCGGC 100085  
 Qy 81 ThrThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAla 100  
 Db 100084 GGTCTTATCAATACAGGAACCGCTGCAACTGCGCGCGCATCGCTTTTGACAAATCG 100025  
 Qy 101 AlaSerLysThr-----HisLeuSerAlaSerIleProAspThrArgMetTrpTyr 118  
 Db 100024 CCCGTCCGCAACGCCGACTACCGCATGAACAGCTACCGACGCGCAACCGCATCTGGTC 99965  
 Qy 119 SerIleGlyAlaThrTyrLysPheThrProAsnLeuSerValAspValGlyPheAlaHis 138  
 Db 99964 TCCCGCGGTATGAATACCATATCGTAAACACCCAGCTGCTCGATGCGCGCTACACCCAC 99905  
 Qy 139 LeuArgGlyLysLysHisPheValGluThrGln----- 150  
 Db 99904 ATC-----CACATCAACGACACCACTACCGCAGCGGCGAAGCGCGC 99860  
 Qy 151 -----AsnIleLysGlyLeuLeuValGluAlaAspTyrThrThrLysAlaThr 167  
 Db 99859 AACGATGGGACACCAAGGC-----CGCTCTCCGCGACGCTTCAAAAACCCAC 99812  
 Qy 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
 Db 99811 GCCGACATCGGCTGCAATACACCTACCAATTC 99776

RESULT 10

AAA49618

ID AAA49618 standard; DNA; 1401 BP.

XX AAA49618;

AC AAA49618;

XX AAA49618;

DT 27-OCT-2000 (first entry)

XX Neisseria meningitidis BASB044 gene #1.  
 DE  
 XX  
 KW Meningitis; microbial disease; upper respiratory tract infection;  
 KW bacteraemia; invasive bacterial disease; BASB044 gene; ds.  
 XX  
 OS Neisseria meningitidis.  
 FH Key Location/Qualifiers  
 FT CDS 1..1401  
 FT /\*tag= a  
 FT /product= "BASB044 protein"  
 FT sig\_peptide 1..72  
 FT /\*tag= b  
 FT mat\_peptide 73..1401  
 FT /\*tag= c  
 XX  
 PN WC200034482-A2.  
 XX  
 PD 15-JUN-2000.  
 XX  
 XX 07-DEC-1999; 99WO-IB02014.  
 XX  
 PR 08-DEC-1998; 98GB-0026979.  
 PR 08-DEC-1998; 98GB-0026980.  
 PR 17-DEC-1998; 98GB-0028015.  
 PR 03-JAN-1999; 99GB-0000090.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX  
 PI Ruelle J, Verlant VGCL;  
 XX  
 XX WPI: 2000-423427/36.  
 DR P-PSDB; AAB01258.  
 XX  
 PT Novel BASB041, 43, 44 and 48 polypeptides of Neisseria meningitidis  
 PT useful for diagnostic, prophylactic and therapeutic purposes against  
 PT microbial diseases comprise a specific amino acid sequence -  
 XX  
 PS Claim 59; Page 163; 171pp; English.  
 XX  
 CC The present sequence is the BASB044 gene from the Neisseria  
 CC meningitidis strain ATCC 13090. The gene and its protein can be used in  
 CC the prevention and treatment of microbial disease such as bacteraemia,  
 CC meningitis and upper respiratory tract infections. They are particularly  
 CC useful for treating bacterial diseases. They can also be used for  
 CC diagnosing these diseases. The protein sequence shows significant  
 CC homology to the long chain fatty acid transport protein FadL of E. coli.  
 XX  
 SQ Sequence 1401 BP; 372 A; 458 C; 336 G; 235 T; 0 other;

Alignment Scores:  
 Pred. No.: 5.54e-08 Length: 1401  
 Score: 143.50 Matches: 45  
 Percent Similarity: 43.23% Conservative: 38  
 Best Local Similarity: 23.44% Mismatches: 70  
 Query Match: 14.99% Indels: 39  
 DB: 21 Gaps: 7

US-09-995-493-52 (1-179) x AAA49618 (1-1401)

Qy 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31  
 Db 865 AAGCCGCGTTAAATCGTTACGCTGAGTCTTTGTCGGTACACGCTATGTACAAAGTG 924  
 Qy 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPhe----- 49  
 Db 925 TCCGATAAAGCCGACCTGTCGGCGGCGGCGCCACGCGCTTCGATAAG 984  
 Qy 50 -----LysGlu-----LeuArgGlyLysTyrGlnAspGlySer 60  
 Db 985 GCGGAAGTGGTTTGAAGAAGAAAAACCGCTCGTCAAGGCCAA----- 1029

QY 61 GlyTyrGluAlaPheThrLysLysGluGluTyrLysAspAsnSerArgPheAlaIleGly 80  
 Db 1030 ---TCGACCGCAGCACCATCACTCCGCCAACACCTACAAAGTTCGGCTTCGGC 1086  
 QY 81 ThrThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAla 100  
 Db 1087 GGTCTTATCAAAATCAGCAACCGCTGCAACTCGCGCGCGCATCGCTTTTGACAAATCG 1146  
 QY 101 AlaSerLysThr-----HisLeuSerAlaSerIleProAspThrAspArgMetTrpTyr 118  
 Db 1147 CCGGTCGCGCAGCCGACTACCGCATGACAGCGCTGCCGCGGACCGGACCGCATCTGGTTC 1206  
 QY 119 SerIleGlyAlaThrTyrLysPheThrProAsnLeuSerValAspValGlyPheAlaHis 138  
 Db 1207 TCCGCGGTGTAATAATACCATATCGGTAAACACACGCTGCTGATCGCGCTTACACCCAC 1266  
 QY 139 LeuArgGlyLysLysHisPheValGluThrGln----- 150  
 Db 1267 ATC-----CACATCAACGACACCATCTACCGCAGCGCAAGCGCAAGCGGC 1311  
 QY 151 -----AsnIleLysGlyLeuLeuValGluAlaAspTyrThrThrLysAlaThr 167  
 Db 1312 AACGATGTGACAGCAAGGCG-----GCGTCTTCGCGCAGCTTTCAAAACAC 1359  
 QY 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
 Db 1360 GCGGACATCATCGCGCTGCAATACACCTACAAATTC 1395  
 RESULT 11  
 AAZ54521  
 ID AAZ54521 standard; DNA; 1395 BP.  
 XX  
 AC AAZ54521;  
 DT  
 DT 21-MAR-2000 (first entry)  
 DE Neisseria meningitidis ORF 989 partial DNA sequence SEQ ID NO:2989.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
 KW antibacterial; gene therapy; ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO9957280-A2.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PF 30-APR-1999; 99WO-US09346.  
 XX  
 PR 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 02-SEP-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX.  
 DR WPI; 2000-062150/05.  
 DR P-PSDB; AAY75759.  
 XX  
 XX Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics  
 XX  
 PS Claim 7; Page 1405-1406; 1453pp; English.

XX  
 CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941  
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent  
 CC PCR primers used in the exemplification of the present inventions. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.  
 XX  
 SQ Sequence 1395 BP; 380 A; 468 C; 320 G; 227 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 7.36e-08 Length: 1395  
 Score: 142.50 Matches: 45  
 Percent Similarity: 42.71% Conservative: 37  
 Best Local Similarity: 23.44% Mismatches: 71  
 Query Match: 14.89% Indels: 39  
 DB: 21 Gaps: 7  
 US-09-995-493-52 (1-179) x AAZ54521 (1-1395)  
 QY 12 LysGlySerLeuThrLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31  
 Db 859 AAAGCCCGCTTAAATCGTTACGCGCGAGTCTTTCGCTACACGCTATGACAAAGT 918  
 QY 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPhe----- 49  
 Db 919 TCCGACAAAGCCGACCTGTTCCGCGACGTAACCTTGGACCGCCACAGCGCTTCGATAAG 978  
 QY 50 -----LysGlu-----LeuArgGlyLysTyrGlnAspGlySer 60  
 Db 979 GCGGAACCTGTTTGAAGAAAGAAAGAAAGCATCGTCAACGGCAA----- 1023  
 QY 61 GlyTyrGluAlaPheThrLysLysGluGluTyrLysAspAsnSerArgPheAlaIleGly 80  
 Db 1024 ---TCGACCGCAGCACCATCACTCCGCCAACCTACCGCATGCAAGTTCGGCTTCGGC 1080  
 QY 81 ThrThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAla 100  
 Db 1081 GGTCTTATCAAAATCAGCAACCGCTGCAACTCGCGCGCGCATCGCTTTTGACAAATCG 1140  
 QY 101 AlaSerLysThr-----HisLeuSerAlaSerIleProAspThrAspArgMetTrpTyr 118  
 Db 1141 CCGGTCGCGCAGCGCAGCTACCGCATGCAAGCTGCGCGGACCGCAACCGCATCTGGTTC 1200  
 QY 119 SerIleGlyAlaThrTyrLysPheThrProAsnLeuSerValAspValGlyPheAlaHis 138  
 Db 1201 TCCGCGGCTGTAATAATACCATATCGGCAAAACACGCTGCTGATCGCGCTTACACCCAC 1260  
 QY 139 LeuArgGlyLysLysHisPheValGluThrGln----- 150  
 Db 1261 ATC-----CACATCAACGACACCATCTACCGCAGCGCAAGCGCAAGCGGC 1305  
 QY 151 -----AsnIleLysGlyLeuLeuValGluAlaAspTyrThrThrLysAlaThr 167  
 Db 1306 AACGATGTGACAGCAAGGCG-----GCGTCTTCGCGCAGCTTTCAAAACAC 1353  
 QY 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
 Db 1354 GCGGACATCATCGCGCTGCAATACACCTACAAATTC 1389  
 RESULT 12  
 ABA97241  
 ID ABA97241 standard; DNA; 1371 BP.  
 XX  
 AC ABA97241;  
 DT  
 DT 21-MAR-2000 (first entry)  
 DE Neisseria meningitidis ORF 989 partial DNA sequence SEQ ID NO:2989.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
 KW antibacterial; gene therapy; ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO9957280-A2.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PF 30-APR-1999; 99WO-US09346.  
 XX  
 PR 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 02-SEP-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX.  
 DR WPI; 2000-062150/05.  
 DR P-PSDB; AAY75759.  
 XX  
 XX Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics  
 XX  
 PS Claim 7; Page 1405-1406; 1453pp; English.





Db 1600 TCCAATGGCTATACCGTGGCTTTGGGACTAAATACAATTTAGG-----GGCTTT 1650





Qy 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeu--- 130  
 Db 1582 GTTGCAACCTGCA-----AGCGTTCGCAACACTTTACCGACCACTGAAA 1629  
 Qy 131 SerValaspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGln 150  
 Db 1630 TCAGCGCGACATC---TTCGATGCCGCCCAATATCGGACATCCGCTTTGTTCCACCAAA 1686  
 Qy 151 ---AsnIleLysGlyLeuLeuValGluAlaAsp 161  
 Db 1687 TTCAACTTCAACGGCAAAACTGGTTCCGTTGAC 1722  
 RESULT 18  
 AAD17037  
 ID AAD17037 standard; DNA; 1941 BP.  
 XX AAD17037;  
 AC AAD17037;  
 DT 29-NOV-2001 (first entry)  
 XX N. meningitidis strain 394/98 ORF46.1 delta G287N2-953 fusion DNA.  
 DE Heterologous expression; Neisserial protein; open reading frame; ORF;  
 KW ORF46.1 delta G287N2-953 fusion protein; ds.  
 XX Neisseria meningitidis.  
 OS  
 XX Key Location/Qualifiers  
 FH 1..1935  
 FT /\*tag= a  
 FT /product= "N. meningitidis New Zealand strain 394/98  
 FT ORF46.1 delta G287N2-953 fusion protein"  
 XX WO200164920-A2.  
 XX 07-SEP-2001.  
 XX 28-FEB-2001; 2001WO-IB00420.  
 XX 28-FEB-2000; 2000GB-0004695.  
 XX 13-NOV-2000; 2000GB-0027675.  
 XX (CHIR-) CHIRON SPA.  
 XX Arico MB, Comanducci M, Galeotti C, Massignani V, Giuliani MM;  
 PI Pizsa M;  
 XX WPI: 2001-557776/62.  
 DR P-PSDB; AAE10018.  
 PT Heterologous expression for the expression of two or more Neisserial  
 PT proteins in fused state  
 XX Example 2; Page 11; 52pp; English.  
 XX The present invention relates to a method for simultaneous heterologous  
 CC expression of two or more Neisserial proteins which are in a fused  
 CC state. The method is useful for simultaneous heterologous expression of  
 CC two or more Neisserial proteins. A protein that may be unstable or  
 CC poorly expressed on its own is assisted by adding a suitable hybrid  
 CC partner and commercial manufacture is simplified only one exproression and  
 CC purification need to be employed in order to produce two separately-  
 CC useful proteins. The present sequence is a DNA encoding  
 CC Neisseria meningitidis (serogroup B, New Zealand strain 394/98) ORF46.1  
 CC (open reading frame) delta G287N2-953 fusion protein.  
 XX  
 SQ Sequence 1941 BP; 557 A; 494 C; 528 G; 362 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0.0462 Length: 1941  
 Score: 98.00 Matches: 51  
 Percent Similarity: 43.60% Conservative: 24

Best Local Similarity: 29.65% Mismatches: 69  
 Query Match: 10.24% Indels: 28  
 DB: 22 Gaps: 11  
 US-09-995-493-52 (1-179) x AAD17037 (1-1941)  
 Qy 3 AsnGlyValLeuGlyProTyrIle-----GlyLysGlySerLeuThrLeuLysLeu--- 19  
 Db 1252 AACGGCTTTAAGGGACATTGGACGGAAATGGCGCGGGATGTTCCGGAGAGTTTTAC 1311  
 Qy 20 ---ProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAlaIleHis 38  
 Db 1312 GCGCCGCGCGGAGAGAAAGTGGCGGA-----AAA 1341  
 Qy 39 TyrSerTyrLysTyrThrGluTrpSerArg-----PheLysGluLeuArgGly---Lys 55  
 Db 1342 TACAGCTATCGCCACACAGATCGGAAAGGGCGGATTCGGGGTGTTCGCGGCAAAAA 1401  
 Qy 56 TyrGlnAspGlySerGlyTyrGluAlaPheThr---LysLysGluGluTyrLysAspAsn 74  
 Db 1402 GAGCAGGATGATCGCGAGGAGGAGGAGCCACCTACAAAGTGGACGAATATCACGCCAAC 1461  
 Qy 75 SerArgPheAlaIle-----GlyThrTyrSerLeuAsnAspAlaLeuThrLeu 91  
 Db 1462 CCGCGTTTCGGCATCGACCATTTCAACACACGACGACGCGCGGTTTTCACGGTCTG 1521  
 Qy 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111  
 Db 1522 ACCGGTTCGTCGAGTTTCGACCAACGCGGCGTAAATCGACATCACCATCCCC 1581  
 Qy 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeu--- 130  
 Db 1582 GTTGCAACCTGCAA-----AGCGTTCGCAACACTTTACCGACCACTGAAA 1629  
 Qy 131 SerValaspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGln 150  
 Db 1630 TCAGCGCGACATC---TTCGATGCCGCCCAATATCGGACATCCGCTTTGTTCCACCAA 1686  
 Qy 151 ---AsnIleLysGlyLeuLeuValGluAlaAsp 161  
 Db 1687 TTCAACTTCAACGGCAAAACTGGTTCCGTTGAC 1722  
 RESULT 19  
 AAZ54328  
 ID AAZ54328 standard; DNA; 2127 BP.  
 XX AAZ54328;  
 AC AAZ54328;  
 XX 21-MAR-2000 (first entry)  
 DT  
 DE  
 XX Neisseria meningitidis ORF 760 partial DNA sequence SEQ ID NO:2605.  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;  
 KW antibacterial; gene therapy; ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO9957280-A2.  
 XX 11-NOV-1999.  
 PD 30-APR-1999; 99WO-US09346.  
 PF 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 02-SEP-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 XX

PA (CHTR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scarlato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX WPI; 2000-062150/05.  
 DR P-PSDB; AAY75566.  
 XX Novel Neisserial polypeptides predicted to be useful antigens for  
 XX vaccines and diagnostics  
 PT Claim 7: Page 1235; 1453pp; English.  
 XX AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941  
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA254573 represent  
 CC PCR primers used in the exemplification of the present inventions. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.  
 XX  
 SQ Sequence 2127 BP; 541 A; 624 C; 554 G; 408 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.081 Length: 2127  
 Score: 96.50 Matches: 44  
 Percent Similarity: 38.3% Conservative: 25  
 Best Local Similarity: 24.4% Mismatches: 70  
 Query Match: 10.08% Indels: 41  
 DB: 21 Gaps: 9

US-09-995-493-52 (1-179) x AA254328 (1-2127)

QY 6 LeuGlyProTyrIleGlyLysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeu 25  
 DB 1690 TTGGGCAACGGCTGTAGTGAAGGTGTGAGACC-----GAAATC 1728  
 QY 26 SerGlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGlu 45  
 DB 1729 ACGGGC-----CGATGACACCGAAGTGGCAATCCATGCAGGTACAGCTACCTGCAC 1782  
 QY 46 TrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPhe 65  
 DB 1783 ---AGCAAAATCAAAACCGCTCAATTCGGCGGACGAAGGC-----ATCTTC 1827  
 QY 66 ThrLysLysGluCluTyrLysAspAsnSerArgPheAlaIleGlyThrThrTyrSerLeu 85  
 DB 1828 CTGCTGATGCCCAACACACACCGCAAC-----CTGTGGACGACTTACCAAGTT 1875  
 QY 86 AsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHis 105  
 DB 1876 ACGTCCGGCTGACCATCGCGCGCGGCGG-----AACCGATGACGGCATTACT 1926  
 QY 106 LeuSerAlaSerIleProAspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLys 125  
 DB 1927 TCATCTGCAGGGATACATACAGCGGCTTATGCCACGCTTCGATCGGCGCATACCGC 1986  
 QY 126 PheThrProAsnLeuSerValaspValGlyPheAlaHisLeuArgGlyLysLysHis 145  
 DB 1987 TTCACGCCCAACTGAAGCTGCAATCAACGCGGACAAATC-----TTCACACGCCCAT 2040  
 QY 146 PheValGluThrGlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThrLys 165  
 DB 2041 TACTAC-----GCGCGGTGCGCAGCGAG 2064

QY 166 AlaThrAlaAsnLeuTyrGly-----LeuAsnLeuAsnTyrArgPhe 179  
 DB 2065 AGCACCTTTAAACATTCCCGGTTCCGAGCGAGCGCTGACGCAACCTGCTTACAGTTT 2124  
 RESULT 20  
 AAA81476  
 ID AAA81476 standard; DNA; 56485 BP.  
 XX AAA81476;  
 AC  
 XX 04-DEC-2000 (first entry)  
 DT  
 XX N. meningitidis partial DNA sequence gnm\_24 SEQ ID NO:24.  
 DE  
 XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB; ds.  
 XX  
 OS Neisseria meningitidis.  
 PN WO200022430-A2.  
 XX  
 PD 20-APR-2000.  
 PF 08-OCT-1999; 99WO-US23573.  
 XX  
 PR 09-OCT-1998; 98US-0103794.  
 PR 30-APR-1995; 99US-0132068.  
 XX  
 PA (CHTR ) CHIRON CORP.  
 XX  
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 PI Rappuoli R, Pizsa M;  
 XX  
 DR WPI; 2000-318079/27.  
 XX  
 PT Isolated nucleotide sequences of Neisseria meningitidis which can be  
 PT used in the diagnosis and treatment of N. meningitidis infection and  
 PT other Neisserial infections, for example, N.gonorrhoea  
 XX  
 PS Claim 7: Page 507-524; 1760pp; English.  
 XX  
 CC The present invention describes methods of obtaining immunogenic  
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
 CC represent specifically claimed Neisseria meningitidis genomic DNA  
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
 CC Neisseria DNA sequences and their corresponding proteins; AAA81294 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to  
 CC Neisserial bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against Meningococcus B; against all serotypes;  
 CC and/or against all pathogenic Neisseriae. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.  
 XX  
 SQ Sequence 56485 BP; 12504 A; 14247 C; 16158 G; 13573 T; 3 other;

Alignment Scores:  
 Pred. No.: 7.93 Length: 56485

Score: 96.50 Matches: 44  
 Percent Similarity: 38.33% Conservative: 25  
 Best Local Similarity: 24.44% Mismatches: 70  
 Query Match: 10.08% Indels: 41  
 DB: 21 Gaps: 9

US-09-995-493-52 (1-179) x AAF21612 (1-349980)

Qy 6 LeuGlyProTyrlleGlyLysGlySerLeuThrLeuLysLeuProAlaTyrlleGlyLeu 25  
 Db 16495 TTGGCAAAACCGGTGATGAGGTGTTGAGACC-----GAAATC 16533

Qy 26 SerGlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyrlleGlySerLeuThrGlu 45  
 Db 16534 AGCGGC-----GGGATGACACCGCAATGCAATCCATGCGAGTTACAGCTACCTGCAC 16587

Qy 46 TrpSerArgPheLysGluLeuArgGlyLysTyrlleGlySerGlyLeuAlaPhe 65  
 Db 16588 ---AGCCAAATCAAAACCGCTCCCAATTCGGCGGACGAGAGC-----ATCTTC 16632

Qy 66 ThrLysLysGluGlyLysAspAsnSerArgPheAlaIleGlyThrTrpSerLeu 85  
 Db 16633 CTGCTGATGCCAACACACAGCGCAAC-----CTGTGGACGACTTACCAAGTT 16680

Qy 86 AsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrlleGlySerLeuThrHis 105  
 Db 16681 ACGTCCGGGCTGACCATCGCGCGCGGTG-----AACGGATGAGCGGCAATTACT 16731

Qy 106 LeuSerAlaSerIleProAspThrAspArgMetTrpTyrlleGlyAlaThrTyrls 125  
 Db 16732 TCATCTCGAGGATACATGACGCGGTATGCCACGTTCGATGCGGCGCATACCGC 16791

Qy 126 PheThrProAsnLeuSerValAspValGlyPheAlaHisLeuArgGlyLysLysHis 145  
 Db 16792 TTCACGCCCAACTGAAGCTGCAATCAACCGCGCACATC-----TTCAACCGCCAT 16845

Qy 146 PheValGluThrGlnAsnIleLysGlyLeuLeuValGluAlaAspTyrlleThrLys 165  
 Db 16846 TACTAC-----GCCCGCGTCCGCGGAG 16869

Qy 166 AlaThrAlaAsnLeuTyrlleGly-----LeuAsnLeuAsnTyrlleArgPhe 179  
 Db 16870 AGCACCTTTAATCTCCGGTTCGGAGCGCAGCTGACGCGCAACCTCGCTTACAGTTT 16929

RESULT 21  
 AAF21612/c  
 ID AAF21612 standard; DNA: 349980 BP.  
 XX  
 AC AAF21612;  
 XX  
 DT 13-MAR-2001 (first entry)  
 XX  
 DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;  
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
 KW ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO200066791-A1.  
 XX  
 PD 09-NOV-2000.  
 XX  
 PF 08-MAR-2000; 2000WO-US05928.  
 XX  
 PR 30-APR-1999; 99US-0132068.  
 PR 08-OCT-1999; 99WO-US23573.  
 PR 28-FEB-2000; 2000GB-0004695.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX

PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Massignani V;  
 PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;  
 PI Frazer CM, Grandi G;  
 XX  
 DR WPI: 2000-647603/62.  
 XX  
 XX Neisseria meningitidis B full length genome sequence and open reading  
 frames are used to detect, treat and prevent Neisserial infections -  
 PT  
 PS Claim 7; Appendix A; 692pp; English.  
 XX  
 XX The present invention describes the full length genome of  
 CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607  
 CC to AAF21613 represent fragments of the NMB genomic sequence, as the  
 CC sequence was too long to go in a record on its own it was split into 8  
 CC sequences which overlap each other at the beginning and end of each  
 CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at  
 CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at  
 CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the  
 CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to  
 CC AAF21606 represent PCR primers which are used in the exemplification of  
 CC the present invention. The NMB genome and fragments from it have  
 CC antibacterial activity, and can be used in vaccines and gene therapy.  
 CC Neisseria nucleic acids, proteins and/or antibodies which binds to the  
 CC proteins can be used in compositions for treating or preventing infection  
 CC due to Neisserial bacteria or as a diagnostic reagent for detecting the  
 CC presence of Neisserial bacteria or of antibodies raised to Neisserial  
 CC bacteria. Computers, computer memory, computer storage medium or computer  
 CC databases can be used in a search to identify open reading frames (ORFs)  
 CC or coding sequences within the NMB genome. The DNA sequences provide  
 CC further opportunities to find antigenic or immunogenic proteins which are  
 CC more effective in vaccines than the outer membrane proteins currently  
 CC used.  
 XX  
 XX Sequence 349980 BP: 86473 A; 95646 C; 85908 G; 81953 T; 0 other;

Alignment Scores:  
 Pred. NO.: 101 Length: 349980  
 Score: 96.50 Matches: 44  
 Percent Similarity: 38.33% Conservative: 25  
 Best Local Similarity: 24.44% Mismatches: 70  
 Query Match: 10.08% Indels: 41  
 DB: 21 Gaps: 9

US-09-995-493-52 (1-179) x AAF21612 (1-349980)

Qy 6 LeuGlyProTyrlleGlyLysGlySerLeuThrLeuLysLeuProAlaTyrlleGlyLeu 25  
 Db 125159 TTGGCAAAACCGGTGATGAGGTGTTGAGACC-----GAAATC 125121

Qy 26 SerGlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyrlleGlySerLeuThrGlu 45  
 Db 125120 AGCGGC-----GGGATGACACCGCAATGCAATCCATGCGAGTTACAGCTACCTGCAC 125067

Qy 46 TrpSerArgPheLysGluLeuArgGlyLysTyrlleGlySerGlyLeuAlaPhe 65  
 Db 125066 ---AGCCAAATCAAAACCGCTCCCAATTCGGCGGACGAGAGC-----ATCTTC 125022

Qy 66 ThrLysLysGluGlyLysAspAsnSerArgPheAlaIleGlyThrTrpSerLeu 85  
 Db 125021 CTGCTGATGCCAACACACAGCGCAAC-----CTGTGGACGACTTACCAAGTT 124974

Qy 86 AsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrlleGlySerLeuThrHis 105  
 Db 124973 ACGTCCGGGCTGACCATCGCGCGCGGTG-----AACGGATGAGCGGCAATTACT 124923

Qy 106 LeuSerAlaSerIleProAspThrAspArgMetTrpTyrlleGlyAlaThrTyrls 125  
 Db 124922 TCATCTCGAGGATACATGACGCGGTATGCCACGTTCGATGCGGCGCATACCGC 124863

Qy 126 PheThrProAsnLeuSerValAspValGlyPheAlaHisLeuArgGlyLysLysHis 145  
 Db 124862 TTCACGCCCAACTGAAGCTGCAATCAACCGCGCACATC-----TTCAACCGCCAT 124809



```
XX 15-JAN-1999; 99GB-0000959.
PR 28-JAN-1999; 99GB-0001903.
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX Ruelle J;
PI WPI: 2000-476062/41.
DR P-PSDB: AAB07697.
XX
XX New Neisseria meningitidis polypeptide useful for diagnosis of
PT Neisseria infection and for development of vaccines against such
PT infection -
XX
XX Claim 11; Page 55; 92pp; English.
XX
XX The present sequence encodes a Neisseria meningitidis BASB053
CC polypeptide. The BASB053 polypeptide, or an antibody immunospecific
CC for BASB053 may be identified in a biological sample in order to
CC diagnose a Neisseria meningitidis infection in an animal. The BASB053
CC polypeptides and polynucleotides may be used as vaccines, for
CC generating an immune response in an animal. A composition comprising
CC at least one antibody immunospecific for BASB053 may be used to
CC treat humans infected with Neisseria meningitidis.
XX
XX Sequence 2169 BP; 540 A; 663 C; 568 G; 398 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 0-111 Length: 2169
Score: 95.50 Matches: 45
Percent Similarity: 36.07% Conservative: 21
Best Local Similarity: 24.59% Mismatches: 70
Query Match: 9.98% Indels: 47
DB: 21 Gaps: 9
US-09-995-493-52 (1-179) x AAS59216 (1-2169)
Oy 6 LeuGlyProTyrIleGlyLysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeu 25
Db 1732 TTGGCAACCGGTGATGGAAGCGTTTCAGACC-----GAAATC 1770
Oy 26 SerGlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGlu 45
Db 1771 AGCGGC-----GCGGTTACACCGAATGGCAATCCATGCAGGTACAGATATCTGCAC 1824
Oy 46 TrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPhe 65
Db 1825 ---AGCCAAATCAAAACCGCTCCCAATTCACGCGACGAGCGC-----ATCTTC 1869
Oy 66 ThrLysLysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThrTyrSerLeu 85
Db 1870 CTGCTGTATGCCCAACACAGCGGCAAC-----CTGTGGACGACTTACCAAGTT 1917
Oy 86 AsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHis 105
Db 1918 ACGCCGAGCTGACCATCGCGCGGAGTG-----AACGCGATGAGCGGCATTACT 1968
Oy 106 LeuSerAlaSerIleProAspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLys 125
Db 1969 TCATCTGCAGGATGCATGCAGCGGGTTATGCCACGTTTCGATGCGATGCGGCATACCGC 2028
Oy 126 PheThrProAsnLeuSerValAspValGly-----Phe 136
Db 2029 TTCACGCCCAAGCTGAGCTGCAATCAACCGCCGACACATCTTCAACCGCATTAATTAC 2088
Oy 137 AlaHisLeuArgGlyLysLysHisPheValGluThrGlnAsnIleLysGlyLeuLeu 156
Db 2089 CCCCCTGTCGCGCGC-----CGGAACACCTTTTACATTTCCCGGT----- 2127
Oy 157 LeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeuLeuTyrGlyLeuAsnLeu 176
Db 2128 -----TCGGAGCGCACCTGGACGCGCAACCTGGGT 2157
```

```
Oy 177 TyrArgPhe 179
Db 2158 TACAGTTTTT 2166
RESULT 24
AAS43869
ID AAS43869 standard; DNA; 1746 BP.
XX AAS43869;
XX
XX 18-DEC-2001 (first entry)
XX Neisseria meningitidis fusion protein delta-G287-953 DNA.
XX Neisseria gonorrhoeae; leader peptide; fusion protein; ORF46.1; ds;
XX Neisseria meningitidis fusion protein.
XX Neisseria meningitidis.
XX Synthetic.
XX WO200164922-A2.
XX 07-SEP-2001.
XX 28-FEB-2001; 2001WO-IB00452.
XX 28-FEB-2000; 2000GB-0004695.
XX 13-NOV-2000; 2000GB-0027675.
XX (CHIR-) CHIRON SPA.
XX Arico MB, Comanducci M, Galeotti C, Masignani V, Gulliani MM;
XX Pizza M;
XX WPI: 2001-582163/65.
XX P-PSDB: AAU27570.
XX Producing heterologous proteins from Neisseria meningitidis and N.
XX gonorrhoeae -
XX Example 15; Page 36-37; 119pp; English.
XX The invention relates to methods for the heterologous expression of
XX Neisseria proteins from Neisseria meningitidis and Neisseria
XX gonorrhoeae. At least one domain in the protein is deleted, e.g. the
XX leader peptide, and may be replaced by a domain from a different protein
XX to make a fusion protein, in order to enhance heterologous expression of
XX Neisseria proteins. Also, a region of a protein, such as a poly-glycine
XX stretch, can be mutated to enhance expression. The proteins used in the
XX processes include ORF46.1, 287, 741, 919, 953, 961 and 983. Sequences
XX AAS43868-AAS43905 represent DNA molecules encoding Neisseria proteins
XX and peptide regions of proteins of the invention.
XX Sequence 1746 BP; 501 A; 449 C; 459 G; 337 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 0.0949 Length: 1746
Score: 95.00 Matches: 50
Percent Similarity: 43.60% Conservative: 25
Best Local Similarity: 29.07% Mismatches: 69
Query Match: 9.93% Indels: 28
DB: 22 Gaps: 11
US-09-995-493-52 (1-179) x AAS43869 (1-1746)
Oy 3 AsnGlyValLeuGlyProTyrIle-----GlyLysGlySerLeuThrLeuLysLeu--- 19
Db 1057 AACGCGCTTAAAGGAGACTTGGACGGAAATGCGCGGGGATGTTCCGGAGAGTTTAC 1116
Oy 20 ---ProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAlaIleHis 38
Db 1117 GGCCCGCGCGGAGAGAGTGGCGGA-----AAA 1146
```





PN WO200228891-A2.  
 XX 11-APR-2002.  
 PD 04-OCT-2001; 2001WO-FR03061.  
 PF 04-OCT-2000; 2000FR-0012697.  
 PR (INSP ) INST PASTEUR.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 PA Kunst F, Glaser P;  
 PI WPI; 2002-332479/37.  
 XX New genomic sequences from *Listeria* species, useful for detection,  
 PT treatment and prevention of infection, also related polypeptides,  
 PT antibodies and modulators  
 XX Claim 16; SEQ ID 918; 180pp; French.  
 PS  
 XX The present invention relates to nucleic acid sequences  
 CC (ABQ671188-ABQ71212) from *Listeria* sp. The sequences are useful as probes  
 CC and primers for identification and/or detection of *Listeria* (e.g. as  
 CC contaminants in foods, or mutational analysis) and for analysis of  
 CC gene expression. Proteins encoded by the nucleic acid sequences can be  
 CC used to screen for compounds that modulate gene expression, replication  
 CC and pathogenicity of *Listeria* (potential therapeutic agents), also for  
 CC treating infections by *Listeria*, and are useful as immunogens in  
 CC anti-*Listeria* vaccines.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1719 BP; 634 A; 286 C; 343 G; 456 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.107 Length: 1719  
 Score: 94.50 Matches: 57  
 Percent Similarity: 39.56% Conservative: 32  
 Best Local Similarity: 25.33% Mismatches: 72  
 Query Match: 9.87% Indels: 64  
 DB: 24 Gaps: 11

US-09-995-493-52 (1-1719) x ABQ68105 (1-1719)

QY 4 GlyValLeuGlyProTyrIleGlyLysGlySerLeuThrLeuLysLeuProAlaTyrTrp 23  
 DB 418 GGAATAAAGGATCATATAATGGCAAA---ACTGTAACAATGAAACCTTGGCAATATAGT 474  
 QY 24 GluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLysTyr 43  
 DB 475 GATTCCAAAGGTTGGTATCAATCAACGCAAACTTTGGCCAAATACCCCATCACAAAGAA 534  
 QY 44 ThrGluTrpSerArgPheLysGlyLeuArg---GlyLysTyrGlnAspGlySerGlyTyr 62  
 DB 535 TCTTTAGAGACAATGCCAAAACCTTAGAATGCCCAAGTTGGGACTCTAGTTATTATAC 594  
 QY 63 GluAlaPheThrLysLys-----GluGluTyrLysAsp-----AsnSer 75  
 DB 595 AAAGGTGCATGCGCGAGAGATCAAAACACATACAAGATGCGACTGCGATGGTTACAGGGA 654  
 QY 76 ArgPheAlaIleGlyThrTyrSer-----LeuAsn----- 86  
 DB 655 COTTATGCAACGGACAACACATATGCTTCTTAAGCTTAATATACGTAATTTCTTCATATAAT 714  
 QY 87 -----Asp 87  
 DB 715 TTGACTCAATATGACTACTCTGTACGATAGGATTAAACAACAAAAAATGTTTCTGAAGAT 774  
 QY 88 AlaLeuThrLeuArgAla-----GlyLeuAlaTyrAspLysAla 100  
 DB 775 GCTAAAGTAGTTAAAGCAGATGGCGATGGTGTGTTTATTAGTGGAAATTTTACAATACGTCCTGCA 834

QY 101 AlaSerLysThrHisLeuSerAlaSerIleProAspThrAspArg----- 115  
 DB 835 GCCAGTGCAGAAAAAGTTATCTACTGCGAGCGCTTACACAACATAAAGATGTAAAAATTTTA 894  
 QY 116 -----MetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSerValAsp 133  
 DB 895 AAAGAAGGCACACTACAAGCAGAGACGGTGGTCCCAATTTTCTCTCAATAATAAGTA--- 951  
 QY 134 ValGlyPheAlaHisLeuArgGly-----LysLysLysHisPheValGluThr 149  
 DB 952 ATCGGTGGATGATGAACACGCCGCAATTTGCTATTATCCAAAGCAACAATGTAAAAACG 1011  
 QY 150 GlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThrLysAlaThrAlaAsn 169  
 DB 1012 CTTAACCTTAACAGGTAATATC-----ACTGCTGGATCTACTAAT 1050  
 QY 170 -LeuTyrGlyLeu 173  
 DB 1051 GGTTTATGCTCTG 1063

RESULT 27  
 ABQ69946  
 ID ABQ69946 standard; DNA; 1719 BP.  
 XX  
 AC ABQ69946;  
 XX  
 DT 29-AUG-2002 (first entry)  
 XX  
 DE *Listeria* monocytogenes EGDe DNA sequence #158.  
 XX  
 KW Antibacterial; *Listeria*; food contamination; mutational analysis;  
 KW infection; ds.  
 XX  
 OS *Listeria* monocytogenes EGDe.  
 XX  
 PN WO200228891-A2.  
 PD 11-APR-2002.  
 XX  
 PF 04-OCT-2001; 2001WO-FR03061.  
 PR 04-OCT-2000; 2000FR-0012697.  
 XX  
 PA (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 PI Kunst F, Glaser P;  
 XX  
 DR WPI; 2002-332479/37.  
 XX  
 PT New genomic sequences from *Listeria* species, useful for detection,  
 PT treatment and prevention of infection, also related polypeptides,  
 PT antibodies and modulators  
 PS Claim 16; SEQ ID 2759; 180pp; French.  
 XX  
 CC The present invention relates to nucleic acid sequences  
 CC (ABQ671188-ABQ71212) from *Listeria* sp. The sequences are useful as probes  
 CC and primers for identification and/or detection of *Listeria* (e.g. as  
 CC contaminants in foods, or mutational analysis) and for analysis of  
 CC gene expression. Proteins encoded by the nucleic acid sequences can be  
 CC used to screen for compounds that modulate gene expression, replication  
 CC and pathogenicity of *Listeria* (potential therapeutic agents), also for  
 CC treating infections by *Listeria*, and are useful as immunogens in  
 CC anti-*Listeria* vaccines.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1719 BP; 634 A; 286 C; 343 G; 456 T; 0 other;

Alignment Scores:



Qy 137 AlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsnIleLysGlyLeuLeu 156  
 Db 1998 GCCCGCTCGCGGC-----GGACACCTTTAATCCTCCCGT----- 2036

Qy 157 LeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeuTyrGlyLeuAsnLeu 176  
 Db 2037 ---TCGGAGCGCAGCTGACGGCA-----AACCTGGCT 2066

Qy 177 TyrArgPhe 179  
 Db 2067 TACACTTTT 2075

RESULT 29  
 ABA91422

ID ABA91422 standard; DNA; 1035 BP.  
 XX  
 AC ABA91422;  
 DT 18-APR-2002 (first entry)  
 DE Haemophilus paragallinarum strain 0222 haemagglutinin haqA gene.  
 KW Haemagglutinin; antigen; vaccine; immunisation; coryza; chicken;  
 KW haqA; gene; ds.  
 XX  
 OS Haemophilus paragallinarum.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1035  
 FT /\*tag= a  
 FT sig\_peptide 1..63  
 FT /\*tag= b  
 FT mat\_peptide 64..1032  
 FT /\*tag= c

WO2000204485-A1.  
 17-JAN-2002.  
 06-JUL-2001; 2001WO-AU00822.  
 07-JUL-2000; 2000AU-0008652.  
 (UYQU ) UNIV QUEENSLAND.  
 Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;  
 WPI; 2002-154917/20.  
 P-PSDB; AAM50721.

New haemagglutinin polypeptide isolated from Haemophilus paragallinarum  
 useful as a vaccine for immunising chickens against coryza caused by  
 the species .

Claim 11; Fig 5; 67pp; English.

The present sequence is that of the coding region of the haqA gene  
 of Haemophilus paragallinarum strain 0222 (serovar B). The gene  
 was isolated from chromosomal DNA by PCR amplification. It  
 encodes haemagglutinin (see AAM50721). The invention provides  
 recombinant haemagglutinin polypeptides (see AAM50716-27) and  
 encoding nucleic acids (see ABA91417-27) from 11 strains (serovars  
 A, B and C) of H. paragallinarum, the causative agent of infectious  
 coryza of chickens. The polypeptides are useful in vaccines for  
 immunisation against infectious coryza, as are the nucleic acids  
 when expressed in attenuated bacteria, especially Salmonella or  
 Mycoplasma (claimed). The recombinant polypeptide is preferably  
 the mature protein, or a biologically active fragment, variant or  
 derivative, that is capable of eliciting an immune response,  
 providing protection against one or more strains of H.  
 paragallinarum in chickens. Also claimed are methods of using the

CC haemagglutinin polypeptides and nucleic acids for detection and  
 CC diagnosis of infectious coryza in chickens.  
 SQ Sequence 1035 BP; 318 A; 186 C; 234 G; 297 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.0816 Length: 1035  
 Score: 93.00 Matches: 38  
 Percent Similarity: 37.50% Conservative: 31  
 Best Local Similarity: 20.65% Mismatches: 71  
 Query Match: 9.72% Indels: 44  
 DB: 24 Gaps: 7

US-09-995-493-52 (1-179) x ABA91422 (1-1035)

Qy 16 ThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrp 35  
 Db 163 ACATTGCGTAATCTGTAACTTATGGGTGTTCCGTTACCAAACTTACTGATAACTTC 222

Qy 36 AlaIleHisTyrSerTyr-----  
 Db 223 GCTGTTGAGCTAGGTTATGATGACTTTGGCGTCTAAACGCGTCAAGCGGTGAAACT 282

Qy 42 -----LysTyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGly 59  
 Db 283 GTTATAAAATACACAAATCACGGAGCTCATTAAAGCTTAAAGCAAGTTATCCAGTGTCT 342

Qy 60 SerGlyTyrGluAlaPheThrLys-----LysGluGluTyrLysAsp 73  
 Db 343 GAAGGATTAGATGTTTATGCTCGCGTGTGAGCAGCGTGTGTTCTGTGATTATAACCA 402

Qy 74 AsnSerArgPheAlaIleGlyThrThr-----TyrSerLeuAsnAspAlaLeuThrLeu 91  
 Db 403 ACTAAAGCAGCAGCTCCTAATCAGACGACGACAACTAGCTTAAAGTTTCTCCAGTATTC 462

Qy 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111  
 Db 463 GCTGCTGCTTAGAGTATAAC-----TTACCATCATTACCA 498

Qy 112 Asp-----ThrAspArgMetTrpTyrSer-----Ile 120  
 Db 499 GAACCTGCATTACGTTGTAATATCAATGGAATAGTAGGACGTGTGGAAGAAAGAT 558

Qy 121 GlyAlaThrTyrLysPheThrProAsnLeu---SerValaspValGlyPheAlaHisLeu 139  
 Db 559 GGTAGCGCTGAGATTATACACCAAGCATCGGTCTCTGTAAGTCTGTGTTATCTTACCGT 618

Qy 140 ArgGlyLysLysLysHisPheValGluThrGlnAsnIleLysGlyLeuLeuValGlu 159  
 Db 619 TTTGTCAAAGTGCACGAGTGTGTAACCTAAGGTTGTTCGCAAAAACATTTTGCATTAAAT 678

Qy 160 AlaAspTyrThr 163  
 Db 679 TCAGATGTTACT 690

RESULT 30  
 ABA91425  
 ID ABA91425 standard; DNA; 1035 BP.  
 XX  
 AC ABA91425;  
 DT 18-APR-2002 (first entry)  
 DE Haemophilus paragallinarum strain H-18 haemagglutinin haqA gene.  
 KW Haemagglutinin; antigen; vaccine; immunisation; coryza; chicken;  
 KW haqA; gene; ds.  
 XX  
 OS Haemophilus paragallinarum.  
 FH Key Location/Qualifiers  
 FT CDS 1..1035  
 FT /\*tag= a

```

FT      /product= "haemagglutinin"
FT      1..63
FT      /tag= b
FT      mat_peptide 64..1032
FT      /tag= c
FT      WO200204485-A1.
PN      17-JAN-2002.
XX
XX      06-JUL-2001; 2001WO-AU00822.
XX
XX      07-JUL-2000; 2000AU-0008652.
XX      (UYQU ) UNIV QUEENSLAND.
XX
XX      Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;
XX      WPI; 2002-154917/20.
XX      P-PSDB; AAM50724.
XX
XX      New haemagglutinin polypeptide isolated from Haemophilus paragallinarum
XX      useful as a vaccine for immunising chickens against coryza caused by
XX      the species
XX
XX      Claim 11; Fig 5; 67pp; English.
XX
XX      The present sequence is that of the coding region of the hga gene
XX      of Haemophilus paragallinarum strain H-18 (serovar C). The gene
XX      was isolated from chromosomal DNA by PCR amplification. It
XX      encodes haemagglutinin (see AAM50724). The invention provides
XX      recombinant haemagglutinin polypeptides (see AAM50716-27) and
XX      encoding nucleic acids (see ABA91417-27) from 11 strains (serovars
XX      A, B and C) of H. paragallinarum, the causative agent of infectious
XX      coryza of chickens. The polypeptides are useful in vaccines for
XX      immunisation against infectious coryza, as are the nucleic acids
XX      when expressed in attenuated bacteria, especially Salmonella or
XX      Mycoplasma (claimed). The recombinant polypeptide is preferably
XX      the mature protein, or a biologically active fragment, variant or
XX      derivative, that is capable of eliciting an immune response,
XX      providing protection against one or more strains of H.
XX      paragallinarum in chickens. Also claimed are methods of using the
XX      haemagglutinin polypeptides and nucleic acids for detection and
XX      diagnosis of infectious coryza in chickens.
XX
XX      Sequence 1035 BP; 318 A; 185 C; 233 G; 299 T; 0 other;
SQ

```

Alignment Scores:

Pred. No.:	0.0816	Length:	1035
Score:	93.00	Matches:	38
Percent Similarity:	37.50%	Conservative:	31
Best Local Similarity:	20.65%	Mismatches:	71
Query Match:	9.72%	Indels:	44
DB:	24	Gaps:	7

US-09-995-493-52 (1-179) x ABA91425 (1-1035)

```

QY      16 ThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrp 35
Db      163 ACATTCGCTAAATCTGTAACTTATGGGTGCTTCGGTGGTACCAATTAAGTAACTTC 222
QY      36 AlaIleHisTyrSerTyr-----LysGluGluTyrLysAsp 41.
Db      223 GCTGTTGAGTAGTTATGATGACTTTGGCGGTGCTAAACTCCGTCAGGCGGTGAAACT 282
QY      42 -----LysTyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGly 59
Db      283 GTTATAAATATACAAATACACGAGCTCACTTAAGCTTAAAGCAAGCTTATCCAGTGCTT 342
QY      60 SerGlyTyrGluAlaPheThrLys-----LysGluGluTyrLysAsp 73
Db      343 GAAGGATTAGATGTTTATGCTCCGCTGGTGGAGCAGCGTTGATTGCTGTGATTATAACCA 402

```

```

QY      74 AsnSerArgPheAlaIleGlyThrThr-----TyrSerLeuAsnAspAlaLeuThrLeu 91
Db      403 ACTAAAGAGCAGCTCTTAATCAGACGACGACAAACATAGCTTAAAGTCTTCTCCAGTATTC 462
QY      92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111
Db      463 GCTGGTGGTGTAGAGTATAAC-----TTACCATCATTTACCA 498
QY      112 Asp-----ThrAspArgMetTrpTyrSer-----Ile 120
Db      499 GAACCTGCATTACCTGTTGAATATCATGCGGTAAATAAGTAGAGCTGTGGAAAAAGAT 558
QY      121 GlyAlaThrTyrLysPheThrProAsnLeu---SerValAspValGlyPheAlaHisLeu 139
Db      559 GGTAGCGGTGATAGATTATATACCAAGCATCGGTTCTGTAACTGCTGGTTTATCTTACCGT 618
QY      140 ArgGlyLysLysHisPheValGluThrGlnAsnIleLysGlyLeuLeuValGlu 159
Db      619 TTTGGTCAAAAGTCACACGAGTTGTTGAACCTTAAGGTTGTTGCCAAAACATTTGCATTAAAT 678
QY      160 AlaAspTyrThr 163
Db      679 TCAGATGTTACT 690
XX
XX      RESULT 31
XX      AAZ54327
XX      ID AAZ54327 standard; DNA; 465 BP.
XX      AC. AAZ54327;
XX      DT 21-MAR-2000 (first entry)
XX      DE Neisseria gonorrhoea ORF 760 partial DNA sequence SEQ ID NO:2603.
XX      KW Neisseria meningitidis; Neisseria gonorrhoea; antigen; vaccine;
XX      KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
XX      KW antibacterial; gene therapy; ds.
XX      OS Neisseria gonorrhoea.
XX      PN WO9957280-A2.
XX      PD 11-NOV-1999.
XX      PF 30-APR-1999; 99WO-US09346.
XX      PR 01-MAY-1998; 98US-0083758.
XX      PR 31-JUL-1998; 98US-0094869.
XX      PR 02-SEP-1998; 98US-0098994.
XX      PR 09-OCT-1998; 98US-0099062.
XX      PR 09-OCT-1998; 98US-0103749.
XX      PR 09-OCT-1998; 98US-0103794.
XX      PR 25-FEB-1999; 99US-0121528.
XX      PA (CHIR ) CHIRON CORP.
XX      PA (GENO-) INST GENOMIC RES.
XX      PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX      PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarfelli M;
XX      PI Tettelin H, Venter JC;
XX      DR WPI; 2000-062150/05.
XX      DR P-PSDB; AAY75565.
XX      PT Novel Neisserial polypeptides predicted to be useful antigens for
XX      PT vaccines and diagnostics
XX      PS Claim 7; Page 1234; 1453pp; English.
XX
XX      AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
XX      represent novel Neisseria meningitidis and N. gonorrhoea polynucleotides
XX      and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ54673 represent

```

CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.

xx  
 SQ Sequence 465 BP; 118 A; 145 C; 119 G; 83 T; 0 other;

## Alignment Scores:

Pred. No.: 0.0308 Length: 465  
 Score: 92.50 Matches: 47  
 Percent Similarity: 36.07% Conservative: 19  
 Best Local Similarity: 25.68% Mismatches: 70  
 Query Match: 9.67% Indels: 47  
 DB: 21 Gaps: 10

US-09-995-493-52 (1-179) x AA254327 (1-465)

Oy 6 LeuGlyProTyrIleGlyLysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeu 25  
 Db 28 TTGGCAAAACCGGTGATGGAAGCGTTGAGACC-----GAAATC 66  
 Oy 26 SerGlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGlu 45  
 Db 67 AGCGGT-----GCGATTACACCGAAATGCCAATCCATGCGAGTTACGATATCGCAC 120  
 Oy 46 TrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPhe 65  
 Db 121 ---AGCCAAATCAAAACCGCGCCCAATCCACGCGACGCGC-----ATCTTC 165  
 Oy 66 ThrLysLysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThrTyrSerLeu 85  
 Db 166 CTGCTGGTGCACCAACACAGCGCAAC-----CTGTGGACGACTTACCAAGTT 213  
 Oy 86 AsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHis 105  
 Db 214 ACGCCCGGGCTGACCGTGGCGGCGCGCTG-----AACGCGATCAGCGCGATTACT 264  
 Oy 106 LeuSerAlaSerIleProAspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLys 125  
 Db 265 TCATCTCGAGGGATGCATCGACGGCGGTATGCCACGTTTCGATGCGATCGCGCATACCGC 324  
 Oy 126 PheThrProAsnLeuSerValAspValGly-----Phe 136  
 Db 325 TTCAGCCCAAGCTGAAGCTGCAATCAATGCCGACACATCTTCAACCGCATTTACTAC 384  
 Oy 137 AlaHisLeuArgGlyLysLysHisPheValGluThrGlnAsnIleLysGlyLeuLeu 156  
 Db 385 GCCCGCGTGGCGGC-----ACGAACACCTTTAATATCCCGGT----- 423  
 Oy 157 LeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeuTyrGlyLeuAsnLeuAsn 176  
 Db 424 ---TCGGAGCGCAGCGTACGCGCA-----AACCTGCGT 453  
 Oy 177 TyrArgPhe 179  
 Db 454 TACAGTTT 462

## RESULT 32

ABA91417

ID ABA91417 standard; DNA; 1035 BP.

xx

AC ABA91417;

xx

DT 18-APR-2002 (first entry)

xx

DE Haemophilus paragallinarum strain 0083 haemagglutinin haqa gene.

xx

KW Haemagglutinin; antigen; vaccine; immunisation; coryza; chicken;  
 KW haqa; gene; ds.

OS Haemophilus paragallinarum.

XX Key Location/Qualifiers

FT CDS 1..1035

FT /tag= a

FT /product= "haemagglutinin"

FT sig\_peptide 1..63

FT /tag= b

FT mat\_peptide 64..1032

FT /tag= c

XX WO200204485-A1.

XX 17-JAN-2002.

XX 06-JUL-2001; 2001WO-AU00822.

XX 07-JUL-2000; 2000AU-0008652.

XX (UYOU ) UNIV QUEENSLAND.

XX Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;

XX WPI; 2002-154917/20.

XX P-PSDB; AAM50716.

XX New haemagglutinin polypeptide isolated from Haemophilus paragallinarum  
 useful as a vaccine for immunising chickens against coryza caused by  
 the species .  
 Claim 11; Fig 5; 67pp; English.

XX The present sequence is that of the coding region of the haqa gene  
 of Haemophilus paragallinarum strain 0083 (serovar A). The gene  
 was isolated from chromosomal DNA by PCR amplification. It  
 encodes haemagglutinin (see AAM50716). The invention provides  
 recombinant haemagglutinin polypeptides (see AAM50716-27) and  
 encoding nucleic acids (see ABA91417-27) from 11 strains (serovars  
 A, B and C) of H. paragallinarum, the causative agent of infectious  
 coryza of chickens. The polypeptides are useful in vaccines for  
 immunisation against infectious coryza, as are the nucleic acids  
 when expressed in attenuated bacteria, especially Salmonella or  
 Mycoplasma (claimed). The recombinant polypeptide is preferably  
 the mature protein, or a biologically active fragment, variant or  
 derivative, that is capable of eliciting an immune response,  
 providing protection against one or more strains of H.  
 paragallinarum in chickens. Also claimed are methods of using the  
 haemagglutinin polypeptides and nucleic acids for detection and  
 diagnosis of infectious coryza in chickens.

XX Sequence 1035 BP; 315 A; 184 C; 235 G; 301 T; 0 other;

## Alignment Scores:

Pred. No.: 0.194 Length: 1035  
 Score: 90.00 Matches: 37  
 Percent Similarity: 36.41% Conservative: 30  
 Best Local Similarity: 20.11% Mismatches: 73  
 Query Match: 9.40% Indels: 44  
 DB: 24 Gaps: 6

US-09-995-493-52 (1-179) x ABA91417 (1-1035)

Oy 16 ThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrp 35

Db 163 ACATTGCGTAATCTGTAACTTATGGGGGTGTTGGTGTATCAAAATCTACTGATAATTTT 222

Oy 36 AlaIleHisTyrSerTyr-----

Db 223 GCTGTTGAGCTAGGTTATGATGACTTTTGGACGTGCGAAACTTCGTCAGACGCGTGAACACT 282

Qy 42 -----LysTyrThrGluTyrSerArgPheLysGluLeuArgGlyLysTyrGlnAspGly 59  
 Db 283 GTTGGAAACATACAAATACGGGGCTCATTTAAGCTTAAAGCAAGTATCCAGTCTT 342  
 Qy 60 SerGlyTyrGluAlaPheThrLys-----LysGluGluTyrLysAsp 73  
 Db 343 GAAGGATTAGATGTTATGTCGGGCTGGAGCTCGTAAATTCGTTCTGATTATAAACCA 402  
 Qy 74 AsnSerArgPheAlaIleGlyThr-----TyrSerLeuAsnAspAlaLeuThrLeu 91  
 Db 403 ACTAAAGACGAGCCCTAATGAGACCGCAGCAACATAGCTTAAAGTTCTCCAGTCTT 462  
 Qy 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111  
 Db 463 GCTGGTGGCTTAGAGTATAAC-----TTACCATCATTAACCA 498  
 Qy 112 AspThrAsp-----ArgMetTyrTyrSerIle 120  
 Db 499 GAATTCGATACGTTGTAATATCAATGGGTGAATAAAGTAGGGCGCTTGGGAAAAAGAT 558  
 Qy 121 GlyAlaThrTyrLysPheThrProAsnLeu---SerValAspValGlyPheAlaHisLeu 139  
 Db 559 GGTAGCCGCTAGATTATACACCAAGCATCGTCTCTGTAACCTGCTGTTTATCTTACCGT 618  
 Qy 140 ArgGlyLysLysLysHisPheValGluThrGlnAsnIleLysGlyLeuLeuValGlu 159  
 Db 619 TTTGGCCAAAGTCACCAAGTGTGTTGAACCTAAGCTTGTGCAAAACATTTGCATTAAAT 678  
 Qy 160 AlaAspTyrThr 163  
 Db 679 TCAGACGTTACT 690

## RESULT 33

ABA91418

ID ABA91418 standard; DNA: 1035 BP.

AC ABA91418;

XX 18-APR-2002 (first entry)

XX Haemophilus paragallinarum strain 221 haemagglutinin *hagA* gene.

XX Haemagglutinin; antigen; vaccine; immunisation; coryza; chicken;

KW *hagA*; gene; ds.

XX Haemophilus paragallinarum.

XX Key Location/Qualifiers

FH CDS 1..1035

FT /\*tag= a

FT /product= "haemagglutinin"

FT sig\_peptide 1..63

FT /\*tag= b

FT mat\_peptide 64..1032

FT /\*tag= c

XX WO200204485-A1.

XX 17-JAN-2002.

XX 06-JUL-2001; 2001WO-AU00822.

XX 07-JUL-2000; 2000AU-0008652.

XX (UYOU ) UNIV QUEENSLAND.

XX Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;

XX WPI: 2002-154917/20.

XX P-PSDB; AAM50717.

XX New haemagglutinin polypeptide isolated from *Haemophilus paragallinarum*  
 PT useful as a vaccine for immunising chickens against coryza caused by

PT the species -

XX Claim 11: Fig 5; 67pp; English.

XX The present sequence is that of the coding region of the *hagA* gene  
 CC of *Haemophilus paragallinarum* strain 221 (serovar A). The gene  
 CC was isolated from chromosomal DNA by PCR amplification. It  
 CC encodes haemagglutinin (see AAM50717). The invention provides  
 CC recombinant haemagglutinin polypeptides (see AAM50716-27) and  
 CC encoding nucleic acids (see ABA91417-27) from 11 strains (serovars  
 CC A, B and C) of *H. paragallinarum*, the causative agent of infectious  
 CC coryza of chickens. The polypeptides are useful in vaccines for  
 CC immunisation against infectious coryza, as are the nucleic acids  
 CC when expressed in attenuated bacteria, especially *Salmonella* or  
 CC *Mycoplasma* (claimed). The recombinant polypeptide is preferably  
 CC the mature protein, or a biologically active fragment, variant or  
 CC derivative, that is capable of eliciting an immune response,  
 CC providing protection against one or more strains of *H.*  
 CC *paragallinarum* in chickens. Also claimed are methods of using the  
 CC haemagglutinin polypeptides and nucleic acids for detection and  
 CC diagnosis of infectious coryza in chickens.

XX SQ Sequence 1035 BP; 315 A; 184 C; 235 G; 301 T; 0 other;

## Alignment Scores:

Pred. No.: 0.194 Length: 1035  
 Score: 90.00 Matches: 37  
 Percent Similarity: 36.41% Conservative: 30  
 Best Local Similarity: 20.11% Mismatches: 73  
 Query Match: 9.40% Indels: 44  
 DB: 24 Gaps: 6

US-09-995-493-52 (1-179) x ABA91418 (1-1035)

Qy 16 ThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrp 35

Db 163 ACATTGGCTAATTCGTGAACCTATATGGGGTGTGGTGGTATCAATTAATGATAATTT 222

Qy 36 AlaIleHisTyrSerTyr----- 41

Db 223 GCTGTTGAGTAGTGTATGATGACTTTGGAGCTGCGAACTTCGTCAGACGGTGAACCT 282

Qy 42 -----LysTyrThrGluTyrSerArgPheLysGluLeuArgGlyLysTyrGlnAspGly 59

Db 283 GTTGGAAACATACAAATACGGGGCTCATTTAAGCTTAAAGCAAGTATCCAGTCTT 342

Qy 60 SerGlyTyrGluAlaPheThrLys-----LysGluGluTyrLysAsp 73

Db 343 GAAGGATTAGATGTTATGCTCGCGTTGGAGCTGCGTAAATTCGTTCTGTATTAAACCA 402

Qy 74 AsnSerArgPheAlaIleGlyThr-----TyrSerLeuAsnAspAlaLeuThrLeu 91

Db 403 ACTAAAGACGAGCCCTAATGAGACCGCAGCAACATAGCTTAAAGTTCTCCAGTCTT 462

Qy 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111

Db 463 GCTGGTGGCTTAGAGTATAAC-----TTACCATCATTAACCA 498

Qy 112 AspThrAsp-----ArgMetTyrTyrSerIle 120

Db 499 GAATTCGATACGTTGTAATATCAATGGGTGAATAAAGTAGGGCGCTTGGGAAAAAGAT 558

Qy 121 GlyAlaThrTyrLysPheThrProAsnLeu---SerValAspValGlyPheAlaHisLeu 139

Db 559 GGTAGCCGCTAGATTATACACCAAGCATCGTCTCTGTAACCTGCTGTTTATCTTACCGT 618

Qy 140 ArgGlyLysLysLysHisPheValGluThrGlnAsnIleLysGlyLeuLeuValGlu 159

Db 619 TTTGGCCAAAGTCACCAAGTGTGTTGAACCTAAGGTGTGTGCAAAACATTTGCATTAAAT 678

Qy 160 AlaAspTyrThr 163

Db 679 TCAGACGTTACT 690

RESULT 34  
 ABA91420  
 ID ABA91420 standard; DNA; 1035 BP.  
 XX  
 AC ABA91420;  
 XX  
 DT 18-APR-2002 (first entry)  
 XX  
 DE Haemophilus paragallinarum strain E-3C haemagglutinin haGA gene.  
 XX  
 KW Haemagglutinin; antigen; vaccine; immunisation; coryza; chicken;  
 KW haGA; gene; ds.  
 XX  
 OS Haemophilus paragallinarum.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1035  
 FT /\*tag= a  
 FT /product= "haemagglutinin"  
 FT sig\_peptide 1..63  
 FT /\*tag= b  
 FT mat\_peptide 64..1032  
 FT /\*tag= c  
 XX  
 PN WO200204485-A1.  
 XX  
 PD 17-JAN-2002.  
 XX  
 PF 06-JUL-2001; 2001WO-AU00822.  
 XX  
 PR 07-JUL-2000; 2000AU-0008652.  
 XX  
 PA (UYQU ) UNIV QUEENSLAND.  
 XX  
 PI Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;  
 XX  
 DR WPI; 2002-154917/20.  
 DR P-PSDB; AAM50719.  
 XX  
 PT New haemagglutinin polypeptide isolated from Haemophilus paragallinarum  
 PT useful as a vaccine for immunising chickens against coryza caused by  
 PT the species  
 XX  
 PS Claim 11; Fig 5; 67pp; English.  
 XX  
 CC The present sequence is that of the coding region of the haGA gene  
 CC of Haemophilus paragallinarum strain E-3C (serovar A). The gene  
 CC was isolated from chromosomal DNA by PCR amplification. It  
 CC encodes haemagglutinin (see AAM50719). The invention provides  
 CC recombinant haemagglutinin polypeptides (see AAM50716-27) and  
 CC encoding nucleic acids (see ABA91417-27) from 11 strains (serovars  
 CC A, B and C) of H. paragallinarum, the causative agent of infectious  
 CC coryza of chickens. The polypeptides are useful in vaccines for  
 CC immunisation against infectious coryza, as are the nucleic acids  
 CC when expressed in attenuated bacteria, especially Salmonella or  
 CC Mycoplasma (claimed). The recombinant polypeptide is preferably  
 CC the mature protein, or a biologically active fragment, variant or  
 CC derivative, that is capable of eliciting an immune response,  
 CC providing protection against one or more strains of H.  
 CC paragallinarum in chickens. Also claimed are methods of using the  
 CC haemagglutinin polypeptides and nucleic acids for detection and  
 CC diagnosis of infectious coryza in chickens.  
 XX  
 SQ Sequence 1035 BP; 315 A; 184 C; 235 G; 301 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0.194 Length: 1035  
 Score: 90.00 Matches: 37  
 Percent Similarity: 36.41% Conservative: 30  
 Best Local Similarity: 20.11% Mismatches: 73  
 Query Match: 9.40% Indels: 44  
 DB: 24 Gaps: 6

US-09-995-493-52 (1-179) x ABA91420 (1-1035)  
 QY 16 ThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrp 35  
 |||||  
 Db 163 ACATGGCGTAATTCGTAACTTATGGGGTGTGGTGTATCAAAATCTACTGATAATTT 222  
 QY 36 AlalleHisTyrSerTyr-----  
 |||||  
 Db 223 GCTGTTGAGCTAGGTTATGATGACTTTGGACGTGCGAAACTTCGTCAGACGGTGAAC 282  
 QY 42 -----LysTyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGly 59  
 |||||  
 Db 283 GTTGGAAAACATACAAATCACGGGCTCATTTAAGCTTTAAAGCAAGTTATCCAGTGCTT 342  
 QY 60 SerGlyTyrGluAlaPheThrLys-----LysGluGluTyrLysasp 73  
 |||||  
 Db 343 GAAGGATTAGATGTTTATGCTCGCGCTGGAGCTGGGTTAATTCGTCTGTATATAACCA 402  
 QY 74 AsnSerArgPheAlaIleGlyThrThr-----TyrSerLeuAsnAspAlaLeuThrLeu 91  
 |||||  
 Db 403 ACTAAACAGCAGCCCTAATGAGACGACGACGACATAGCTTAAAGTTTCTCCAGTGTTT 462  
 QY 92 ArgAlaGlyLeuAlaTyrAspLysAlaSerLysThrHisLeuSerAlaSerIlePro 111  
 |||||  
 Db 463 GCTGGTGGCTTAGAGTATAAC-----TTACCATCATTACCA 498  
 QY 112 AspThrAsp-----ArgMetTrpTyrSerIle 120  
 |||||  
 Db 499 GAACCTGCTTACGTGTTGAATATCAATGGGTGAATAAGTAGGCGTTGGAAAAAGAT 558  
 QY 121 GlyAlaThrTyrLysPheThrProAsnLeu---SerValaspValGlyPheAlaHisLeu 139  
 |||||  
 Db 559 GGTAGCCGTGTAGATTATACACCAAGCATCGTCTCTGTAACTGCTGGTTTATCTTACCGT 618  
 QY 140 ArgGlyLysLysLysHisPheValGluThrGlnAsnIleLysGlyLeuLeuValGlu 159  
 |||||  
 Db 619 TTGGCCAAAGTGCACCATGTTGTTGAACCTAAGTGTGTGCAAAACATTTGCATTAAAT 678  
 QY 160 AlaAspTyrThr 163  
 |||||  
 Db 679 TCAGACGTTACT 690  
 RESULT 35  
 ABA91421  
 ID ABA91421 standard; DNA; 1026 BP.  
 XX  
 AC ABA91421;  
 XX  
 DT 18-APR-2002 (first entry)  
 DE Haemophilus paragallinarum strain HP14 haemagglutinin haGA gene.  
 XX  
 KW Haemagglutinin; antigen; vaccine; immunisation; coryza; chicken;  
 KW haGA; gene; ds.  
 XX  
 OS Haemophilus paragallinarum.  
 FH Key Location/Qualifiers  
 FT CDS 1..1026  
 FT /\*tag= a  
 FT /product= "haemagglutinin"  
 FT sig\_peptide 1..63  
 FT /\*tag= b  
 FT mat\_peptide 64..1023  
 FT /\*tag= c  
 XX  
 PN WO200204485-A1.  
 XX  
 PD 17-JAN-2002.  
 XX  
 PF 06-JUL-2001; 2001WO-AU00822.  
 XX





```

CC diagnosis of infectious coryza in chickens.
XX
SQ Sequence 1026 BP; 316 A; 188 C; 228 G; 294 T; 0 other;

Alignment Scores:
Pred. No.: 0.222 Length: 1026
Score: 89.50 Matches: 37
Percent Similarity: 38.12% Conservative: 32
Best Local Similarity: 20.44% Mismatches: 71
Query Match: 9.35% Indels: 41
DB: 24 Gaps: 7

US-09-995-493-52 (1-179) x ABA91427 (1-1026)
Qy 16 ThrLeuLysLeuProAlaTyrTrpCluLeuSerGlyPheHisGlnLeuThrAspGlnTrp 35
Db 163 ACATTGCGTAATCTGTAACCTATGCTGTTGGTGTACCAAAATACGATAACTTC 222
Qy 36 AlaIleHisTyrSerTyr-----
Db 223 GCTGTTGAGCTAGCTTATGATGACTTTGGAGCTGCGAAATTCGCCAAGACGGTGAAACT 282
Qy 42 -----LysTyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGly 59
Db 283 GTTATAAACACATACAAATCAGCGGCTCATTAAAGCAAGTTATCCAGTCGCTT 342
Qy 60 SerGlyTyrGluAlaPheThrLys-----LysGluGluTyrLysAsp 73
Db 343 GAGGGTTAGATGTTATGCTGCGGTTGGACGACGATGATTCGTTCTGTGATTATAACCA 402
Qy 74 AsnSerArgPheAlaIleGlyThr-----TyrSerLeuAsnAspAlaLeuThrLeu 91
Db 403 ACTAAAGACGAGCGCCCTAATGAGACGACGACGACGCTTAAAGTATCTCCAGTATTT 462
Qy 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlepro 111
Db 463 GCAGAGGTTTAGAGTATAAC-----TTACCATCATTTACCA 498
Qy 112 Asp-----ThrAspArgMetTrpTyrSer-----IleGlyAlaThr 123
Db 499 GAACCTGCACTACGTTGTAATATCAATGCGTAATAAGTAGGACGTTAGCGGT 558
Qy 124 TyrLysPheThrProAsnLeu---SerValAspValGlyPheAlaHisLeuArgGlyLys 142
Db 559 CTAGATTATACCAACGACGCTGCTGTAACCTGCTGTTTATCTCTACCGGTTTGGTCAA 618
Qy 143 LysLysHisPheValGluThrGlnAsnIleLysGlyLeuLeuValGluAlaAspTyr 162
Db 619 AGTCACCACTGTTGTAACCTAAGGTTGTCGCAAAACATTTGCCATTAAATTCAGATGTT 678
Qy 163 Thr 163
Db 679 ACT 681

RESULT 37
ABA91427
ID ABA91427 standard; DNA; 1026 BP.
XX
AC ABA91427;
XX
XX 18-APR-2002 (first entry)
XX
DE Haemophilus paragallinarum strain HP60 haemagglutinin in hgaA gene.
XX
KW Haemagglutinin; antigen; vaccine; immunisation; coryza; chicken;
KW hgaA; gene; ds.
XX
OS Haemophilus paragallinarum.
XX
FH Key Location/Qualifiers
FT CDS 1..1026
FT /tag= a
FT /product= "haemagglutinin"

```

```

FT sig_peptide 1..63
FT mat_peptide /*tag= b
FT 64..1023 /*tag= c
XX
PN WO200204485-A1.
XX
XX 17-JAN-2002.
XX
XX 06-JUL-2001; 2001WO-AU00822.
XX
XX 07-JUL-2000; 2000AU-0008652.
XX
XX (UYQU ) UNIV QUEENSLAND.
XX
XX Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;
XX
XX WPI; 2002-154917/20.
XX
XX P-PSDB; AAM50726.
XX
XX New haemagglutinin polypeptide isolated from Haemophilus paragallinarum
XX useful as a vaccine for immunising chickens against coryza caused by
XX the species
XX
XX Claim 11; Fig 5; 67pp; English.
XX
XX The present sequence is that of the coding region of the hgaA gene
XX of Haemophilus paragallinarum strain HP60 (serovar C). The gene
XX was isolated from chromosomal DNA by PCR amplification. It
XX encodes haemagglutinin (see AAM50726). The invention provides
XX recombinant haemagglutinin polypeptides (see AAM50716-27) and
XX encoding nucleic acids (see ABA91417-27) from 11 strains (serovars
XX A, B and C) of H. paragallinarum, the causative agent of infectious
XX coryza of chickens. The polypeptides are useful in vaccines for
XX immunisation against infectious coryza, as are the nucleic acids
XX when expressed in attenuated bacteria, especially Salmonella or
XX Mycoplasma (claimed). The recombinant polypeptide is preferably
XX the mature protein, or a biologically active fragment, variant or
XX derivative, that is capable of eliciting an immune response,
XX providing protection against one or more strains of H.
XX paragallinarum in chickens. Also claimed are methods of using the
XX haemagglutinin polypeptides and nucleic acids for detection and
XX diagnosis of infectious coryza in chickens.
XX
SQ Sequence 1026 BP; 316 A; 188 C; 228 G; 294 T; 0 other;

Alignment Scores:
Pred. No.: 0.222 Length: 1026
Score: 89.50 Matches: 37
Percent Similarity: 38.12% Conservative: 32
Best Local Similarity: 20.44% Mismatches: 71
Query Match: 9.35% Indels: 41
DB: 24 Gaps: 7

US-09-995-493-52 (1-179) x ABA91427 (1-1026)
Qy 16 ThrLeuLysLeuProAlaTyrTrpCluLeuSerGlyPheHisGlnLeuThrAspGlnTrp 35
Db 163 ACATTGCGTAATCTGTAACCTATGCTGTTGGTGTACCAAAATACGATAACTTC 222
Qy 36 AlaIleHisTyrSerTyr-----
Db 223 GCTGTTGAGCTAGCTTATGATGACTTTGGAGCTGCGAAATTCGCCAAGACGGTGAAACT 282
Qy 42 -----LysTyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGly 59
Db 283 GTTATAAACACATACAAATCAGCGGCTCATTAAAGCAAGTTATCCAGTCGCTT 342
Qy 60 SerGlyTyrGluAlaPheThrLys-----LysGluGluTyrLysAsp 73
Db 343 GAGGGTTAGATGTTATGCTGCGGTTGGACGACGATGATTCGTTCTGTGATTATAACCA 402
Qy 74 AsnSerArgPheAlaIleGlyThr-----TyrSerLeuAsnAspAlaLeuThrLeu 91
Db 403 ACTAAAGACGAGCGCCCTAATGAGACGACGACGACGCTTAAAGTATCTCCAGTATTT 462
Qy 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlepro 111
Db 463 GCAGAGGTTTAGAGTATAAC-----TTACCATCATTTACCA 498
Qy 112 Asp-----ThrAspArgMetTrpTyrSer-----IleGlyAlaThr 123
Db 499 GAACCTGCACTACGTTGTAATATCAATGCGTAATAAGTAGGACGTTAGCGGT 558
Qy 124 TyrLysPheThrProAsnLeu---SerValAspValGlyPheAlaHisLeuArgGlyLys 142
Db 559 CTAGATTATACCAACGACGCTGCTGTAACCTGCTGTTTATCTCTACCGGTTTGGTCAA 618
Qy 143 LysLysHisPheValGluThrGlnAsnIleLysGlyLeuLeuValGluAlaAspTyr 162
Db 619 AGTCACCACTGTTGTAACCTAAGGTTGTCGCAAAACATTTGCCATTAAATTCAGATGTT 678
Qy 163 Thr 163
Db 679 ACT 681

RESULT 37
ABA91427
ID ABA91427 standard; DNA; 1026 BP.
XX
AC ABA91427;
XX
XX 18-APR-2002 (first entry)
XX
DE Haemophilus paragallinarum strain HP60 haemagglutinin in hgaA gene.
XX
KW Haemagglutinin; antigen; vaccine; immunisation; coryza; chicken;
KW hgaA; gene; ds.
XX
OS Haemophilus paragallinarum.
XX
FH Key Location/Qualifiers
FT CDS 1..1026
FT /tag= a
FT /product= "haemagglutinin"

```





CC The present invention relates to a Moraxella catarrhalis genomic library  
CC comprising of a combination of 41 nucleic acid molecules (see  
CC AAF28514-AAF28554). The library has a number of uses described in the  
CC specification e.g. is useful for identifying diagnostic and therapeutic  
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large  
CC aerobic, gram-negative diplococcus, normally found among the bacterial  
CC flora of human upper airways. M. catarrhalis is known to cause acute,  
CC localised infections such as otitis media, sinusitis and bronchopulmonary  
CC infection and life-threatening, systemic diseases including endocarditis  
CC and meningitis.

XX  
SQ Sequence 119211 BP; 34729 A; 23773 C; 25664 G; 35041 T; 4 other;

## Alignment Scores:

Pred. No.:	198	Length:	119211
Score:	89.00	Matches:	42
Percent Similarity:	33.70%	Conservative:	19
Best Local Similarity:	23.20%	Mismatches:	84
Query Match:	9.30%	Indels:	36
DB:	22	Gaps:	6

US-09-995-493-52 (1-179) x AAF28553 (1-119211)

QY	9	TyrIleGlyLysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPhe	28
DB	25749	TATGCGGGTAAAGAGATTAAAGTTACTTTGCCCTGACTCATGGAACCTTAGATTTC	25808
QY	29	HisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArg	48
DB	25809	ACTGGTGTGAATCCCAACACGGTATTACTGCCAAGTACGCTATGTACCATGGTCTGAT	25868
QY	49	PheLysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLys	68
DB	25869	TTTGACATTCCGCCCAACACAGTATACAGAA-----ACCACAAAA	25907
QY	69	GluGluTyr-----LysAspAsnSerArgPheAla	78
DB	25908	CTTCGTTATCCACAGGGTTTACCAATCATCATGACAAAGACCAATGGTCGGCTGAA	25967
QY	79	IleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAsp	98
DB	25968	GTTGGTTGGGTAACGGTGTAGCGATCGTTTGGCTGTTTCAGGTCCGGTAGTGGGAT	26027
QY	99	LysAlaLeuSerLysThrHisLeuSerAlaSerIleProAspThrAspArgMetTrpTyr	118
DB	26028	ACTGGTGCAGGTAAC-----CCTGCAAGTAGCTTAGGTCTATCAAGGCTATTAT	26078
QY	119	Ser-----IleGlyAlaThrTyrLysPheThrProAsnLeuSerValAspValGlyPhe	136
DB	26079	TCATTGGGCTTAGGTCCCGTTATATGTTACACCTGAATGTCGCTGCTTTGGGT---	26135
QY	137	AlaHisLeuArgGlyLysLysHisPheValGluThrGlnAsnIleLysGlyLeuLeu	156
DB	26136	-----GGTAATACTTTAAATTTGGAGATGCTCAA-----	26165
QY	157	LeuValGluAlaAspTyrThrLysAlaThrAlaAsnLeuTyrGlyLeuAsnLeuAsn	176
DB	26166	-----GCACAGCTACCAACCAAGATAAAGTAGGTAACCTTTTGATAGTATGATGC	26216
QY	177	Tyr	177
DB	26217	TAT	26219

Search completed: May 19, 2003, 19:21:57  
Job time : 902 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 19, 2003, 19:13:38 : Search time 42 Seconds  
(without alignments)  
1307.027 Million cell updates/sec

Title: US-09-995-493-52  
Perfect score: 957  
Sequence: 1 QHNGVLGPIYIGKSLTKLP.....ADYTTKATNLYGLNLYRF 179

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US0995493/runat\_12052003\_091309\_23030/app\_query.fasta\_1.327  
-DB-Issued\_Patents\_NA -QFMT=fastap -SUFFIX=run1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0995493.ecgn\_1.1.32.0runat\_12052003\_091309\_23030 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-FGAPOP=10 -FGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	570.5	59.6	1347	2	US-08-472-172-5
2	559	58.4	1365	2	US-08-472-172-3
3	549	57.4	1381	4	US-08-476-102A-5
4	549	57.4	1560	2	US-08-472-172-1
5	549	57.4	1598	2	US-08-743-637B-26
6	549	57.4	1598	3	US-08-526-840B-26
7	256	26.8	240	2	US-08-472-172-11
8	89	9.3	1650	1	US-08-245-758-11
9	89	9.3	1650	2	US-08-810-655A-1
10	89	9.3	1650	5	PCT-US95-05134-11
11	87	9.1	23451	4	US-09-453-702B-173
12	83.5	8.7	784	3	US-08-961-083-151

13	83.5	8.7	1245	1	US-08-183-214-1	Sequence 1, Appl1
14	83.5	8.7	2903	4	US-09-221-017B-35	Sequence 35, Appl
c 15	83.5	8.7	11887	4	US-08-961-527-146	Sequence 146, App
c 16	83	8.7	9057	4	US-09-453-702B-194	Sequence 194, App
17	81.5	8.5	1103	1	US-08-183-214-9	Sequence 9, Appl1
18	81	8.5	2319	1	US-08-286-767-1	Sequence 1, Appl1
19	81	8.5	2320	1	US-08-090-526-1	Sequence 1, Appl1
20	81	8.5	2320	1	US-08-486-715-1	Sequence 1, Appl1
21	81	8.5	2320	1	US-08-486-719-1	Sequence 1, Appl1
22	81	8.5	2320	1	US-08-271-354-1	Sequence 1, Appl1
23	81	8.5	2320	1	US-08-476-100-1	Sequence 1, Appl1
24	81	8.5	2320	2	US-08-565-861-1	Sequence 1, Appl1
25	81	8.5	2320	3	US-08-475-749-1	Sequence 1, Appl1
26	81	8.5	2320	5	PCT-US94-07658-1	Sequence 1, Appl1
27	80	8.4	547	4	US-09-370-838-169	Sequence 169, App
28	80	8.4	2097	4	US-08-848-760B-25	Sequence 25, Appl
29	80	8.4	2327	2	US-08-175-158A-1	Sequence 1, Appl1
30	80	8.4	3427	1	US-08-470-058-1	Sequence 1, Appl1
31	80	8.4	3427	3	US-09-037-188-1	Sequence 1, Appl1
32	80	8.4	3427	4	US-09-285-310-1	Sequence 1, Appl1
33	80	8.4	4601	1	US-08-470-058-3	Sequence 3, Appl1
34	80	8.4	4601	3	US-09-037-188-3	Sequence 3, Appl1
35	80	8.4	4601	4	US-09-285-310-3	Sequence 3, Appl1
36	79	8.3	6519	1	US-08-233-008A-7	Sequence 7, Appl1
37	78	8.2	2220	5	PCT-US95-13749-2	Sequence 2, Appl1
38	77.5	8.1	2600	4	US-08-628-434-1	Sequence 1, Appl1
c 39	77	8.0	852	4	US-09-277-565-24	Sequence 24, Appl
40	76	7.9	3074	4	US-09-087-277-1	Sequence 1, Appl1
41	76	7.9	3074	4	US-09-658-499-1	Sequence 1, Appl1
42	75.5	7.9	1109	1	US-08-183-214-11	Sequence 11, Appl
43	74.5	7.8	10711	4	US-08-961-527-145	Sequence 145, App
44	74	7.7	849	3	US-08-733-230-5	Sequence 5, Appl1
45	74	7.7	849	4	US-08-953-326-5	Sequence 5, Appl1

ALIGNMENTS

RESULT 1  
US-08-472-172-5  
: Sequence 5, Application US/08472172  
: Patent No. 5985288  
: GENERAL INFORMATION:  
: APPLICANT: Munson, Jr., Robert S  
: APPLICANT: Grass, Susan  
: APPLICANT: Chong, Pele Y Y  
: APPLICANT: Fahim, Raafat  
: APPLICANT: Sia, Charles D Y  
: APPLICANT: McVerfry, Patrick  
: APPLICANT: Klein, Michel  
: TITLE OF INVENTION: Outer Membrane Protein p1 and Peptides  
: of Haemophilus Influenzae Type B  
: NUMBER OF SEQUENCES: 26  
: CORRESPONDENCE ADDRESSES:  
: ADDRESSEE: Sim & McBurney  
: STREET: Suite 701, 330 University Avenue  
: CITY: Toronto  
: STATE: Ontario  
: COUNTRY: Canada  
: ZIP: M5G 1R6  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/472,172  
: FILING DATE: 07-JUN-1995  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/849,411  
: FILING DATE: 07-JUL-1992  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:

```

; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-471
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; TELEX: 063-24567 SIMBAS
; INFORMATION FOR SEQ ID NO: 5:
; LENGTH: 1347 base pairs
; SEQUENCE CHARACTERISTICS:
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-472-172-5

```

```

Alignment Scores:
Pred. No.: 2,89e-64 Length: 1347
Score: 570.50 Matches: 106
Percent Similarity: 80.47% Conservative: 30
Best Local Similarity: 62.72% Mismatches: 30
Query Match: 59.61% Indels: 3
DB: 2 Gaps: 2

```

```
US-09-995-493-52 (1-179) x US-08-472-172-5 (1-1347)
```

```

Qy 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 847 AAAGGTAATTAACCCCTAAATGCCAGATTACTTAGAACCTTCTGGTTCCATCAATTA 906
Qy 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 907 ACTGACAAATTTGCTGTCATATAGTTATATAATATACCCCATTTGGAGTCGTTAACAAAA 966
Qy 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 967 TTACATGCTAGCTCGAAGATGGTAAG-----AAAGCTTTTGATAAAGATTGCAATAC 1020
Qy 72 LysAspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeu 91
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1021 AGTAATAACTCTCGTATGTCATTAGGCGCAAGTTATATCTTGATGAAAAATTGACCTTA 1080
Qy 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1081 CGTCGGGTATGCTTACGATCAAGCTGCATCTCGTCATACCCCTAGTCGTGCAATTCCTCA 1140
Qy 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1141 GATACCAATCGCACTTGGTATAGTTTAGGGGCAACCTATATAATTCACGCCCAATTTATCT 1200
Qy 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1201 GTTCATCTTGGCTATGCTTACTTAAAGGCAAAAAGTTCACTTTAAAGAGCACACAA 1260
Qy 152 IleLys---GlyLeuLeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeu 170
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1261 GCTCAGGTGGCTTCATACACAAACCGCAATTTACACTTCTCAAGCACACGCAATCTT 1320
Qy 171 TyrGlyLeuAsnLeuAsnTyrArgPhe 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1321 TACGGCTTAAACTTAAATATATAGTTTC 1347

```

# RESULT 2

```

US-08-472-172-3
; Sequence 3, Application US/08472172
; Patent No. 5985288
; GENERAL INFORMATION:
; APPLICANT: Munson, Jr., Robert S
; APPLICANT: Grass, Susan
; APPLICANT: Chong, Pele Y
; APPLICANT: Fahim, Raafat
; APPLICANT: Sia, Charles D Y
; APPLICANT: McVerry, Patrick
; APPLICANT: Klein, Michel

```

```

; TITLE OF INVENTION: Outer Membrane Protein P1 and Peptides
; TITLE OF INVENTION: of Haemophilus Influenzae Type B
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R6
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,172
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/849,411
; FILING DATE: 07-JUL-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-471
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; TELEX: 063-24567 SIMBAS
; INFORMATION FOR SEQ ID NO: 3:
; LENGTH: 1365 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-472-172-3

```

```

Alignment Scores:
Pred. No.: 9,02e-63 Length: 1365
Score: 559.00 Matches: 106
Percent Similarity: 78.49% Conservative: 29
Best Local Similarity: 61.63% Mismatches: 31
Query Match: 58.41% Indels: 6
DB: 2 Gaps: 2

```

```
US-09-995-493-52 (1-179) x US-08-472-172-3 (1-1365)
```

```

Qy 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 856 AAAGGTAATTAACCCCTTACATGCCAGATTACTTAGAACCTTCTGGGTCCATCAATTA 915
Qy 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 916 ACTGACAAATTTGCGTGCATATAGTTATATAATATACCCCATTTGGAGTCGTTAACAAAA 975
Qy 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 976 TTACACGCTAGCTTCGAAGATGGTAAG-----AAAGCTTTTGATAAAGAAATTAACATAC 1029
Qy 72 LysAspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeu 91
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1030 AGTAATAACTCTCGTGTGTCATTAGGCGCAAGTTATATCTTGATGAAAAATTAACCTTA 1089
Qy 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1090 CGTCGGGTATGCTTACGATCAAGCTGCATCTCGTCATACCCGTAGTCGTGCAATTCCTCA 1149
Qy 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1150 GATACCGATCGCACTTGGTATAGTTTAGGTGCAACCTATATAATTTACGCCCAATTTATCT 1209

```



Qy 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysLysHisPheValGluThrGlnAsn 151  
Db 1210 GTTGATCTGGCTATGCTTACTTAAAGGCAAAAGTTCACTTTAAAGAGTAAACAA 1269  
Qy 152 IleLysGly-----LeuLeuLeuValGluAlaAspTyrThrThrLysAlaThr 167  
Db 1270 ATAGGTGCAACAGGTACATTGACATTGAATACAACTGCAAAATATATCTCTCAAGCACAC 1329  
Qy 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
Db 1330 GCAATCTTTACGGTTTGAATTAATTAATATAGTTTC 1365

## RESULT 3

US-08-476-102A-5  
; Sequence 5, Application US/08476102A  
; Patent No. 6355450  
; GENERAL INFORMATION:  
; APPLICANT: Fleischmann, et al.  
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus  
; Patent No. 6355450

; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Ave.  
; CITY: Rockville  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20850

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/476,102A  
FILING DATE: 07-Jun-1995  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Marks, Michelle S.  
REGISTRATION NUMBER: 41,971  
REFERENCE/DOCKET NUMBER: PB186P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-309-8504  
TELEX: 301-309-8439  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1381 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:

## US-08-476-102A-5

Alignment Scores:  
Pred. No.: 1.8e-61 Length: 1381  
Score: 549.00 Matches: 106  
Percent Similarity: 77.91% Conservative: 28  
Best Local Similarity: 61.63% Mismatches: 32  
Query Match: 57.37% Indels: 6  
DB: 4 Gaps: 3

## US-08-476-102A-5

US-08-476-102A-5 (1-179) x US-08-476-102A-5 (1-1381)  
Qy 12 LysGlySerLeuLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31  
Db 872 AAAGTGATTAACCTTACATTGCGACATTACTTAGAACCTTCGGTTCCATCAATTA 931  
Qy 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51  
Db 932 ACTGACAACTTGTGTCGATTATAGTTATATATATATATATATATATATATATATAT 991

Qy 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGlyTyr 71  
Db 992 TTAACGCGTAGCTCGAAGATGATAA-----AAAGCTTTTGTATAAAGAAATTACAATAC 1045  
Qy 72 LysAspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeu 91  
Db 1046 AGTAATAACTCTCGTGTTCATTAGGGCAAGTTATAATCTTGTATGATGAAAAATTTGACCTTA 1105  
Qy 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIlePro 111  
Db 1106 CGTGGGGGTATGCTTACGATCAGGCTGCATCTGCATCAGCGTAGTGCTGCAATTCGA 1165  
Qy 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131  
Db 1166 GATACCGATCGCACCTTGGTATAGTTTAGGGCAACCTATAAATTCAGCGCAATTTATCT 1225  
Qy 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151  
Db 1226 GTTGATCTTGGCTATGCTTACTTAAAGGCAAAAGTTCACTTTAAAGAAAGTAAACAA 1285  
Qy 152 Ile-----LysGlyLeuLeuLeu-----ValGluAlaAspTyrThrThrLysAlaThr 167  
Db 1286 ATAGGTGACGAACGTTCACTTACATTGAATCAACACTGCAAAATATATATCTCTCAAGCACAT 1345  
Qy 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
Db 1346 GCAATCTTTACGGTTTAAACTTAAATTAATATATAGTTTC 1381

## RESULT 4

US-08-472-172-1  
; Sequence 1, Application US/08472172  
; Patent No. 5985288  
; GENERAL INFORMATION:  
; APPLICANT: Munson, Jr., Robert S  
; APPLICANT: Grass, Susan  
; APPLICANT: Chong, Pele Y Y  
; APPLICANT: Fahim, Raafat  
; APPLICANT: Sia, Charles D Y  
; APPLICANT: McVerry, Patrick  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Outer Membrane Protein P1 and Peptides  
; TITLE OF INVENTION: Of Haemophilus Influenzae Type B  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R6  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/472,172  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/849,411  
; FILING DATE: 07-JUL-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-471  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEX: (416) 595-1163  
; TELETYPE: 063-24567 SIMBAS  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:



```
Db 1413 GCAATCTTTACGGTTTGAATTAATATAGTTTC 1448
|||||
RESULT 6
US-08-526-840B-26
; Sequence 26, Application US/08526840B
; Patent No. 6001564
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,840B
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/304,732
; FILING DATE: 12-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1598 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Haemophilus influenzae
US-08-526-840B-26
Alignment Scores:
Pred. No.: 2,21e-61 Length: 1598
Score: 549.00 Matches: 105
Percent Similarity: 76.74% Conservative: 27
Best Local Similarity: 61.05% Mismatches: 34
Query Match: 57.37% Indels: 6
DB: 3 Gaps: 2
US-09-995-493-52 (1-179) x US-08-526-840B-26 (1-1598)
Qy 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31
|||||
Db 939 AAGGTAATTTAACTTTACATGGCAGATTACTTAGAATCTTCGGTTCCATCAATTA 998
|||||
Qy 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51
|||||
Db 999 ACTGACAACTTGCAGTCGATTATAGTTATAATATACCATTCGAGTCGTTTACAAAAA 1058
|||||
Qy 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71
|||||
|||||
```

```
Db 1059 TTACATGCCAGCTTCGAGATGGTAAA-----AAAGCTTTTCATAAAGAATTACAATAC 1112
Qy 72 LysAspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeu 91
|||||
Db 1113 AGTAATAACTCTCGTTCATAGGGCAAGTTATAATCTTTATGAAAAATTGACCTTA 1172
|||||
Qy 92 ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLysLeuSerAlaSerIlePro 111
|||||
Db 1173 COTGGCGGTATTGCTTACCATCAAGCGCATCTCGTCATCACCGTAGTCTGCAATTCCA 1232
|||||
Qy 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131
|||||
Db 1233 GATACGATCGCATCTGGTATAGTTAGTGCAACCTATAAATTCACGCGGAATTTATCT 1292
|||||
Qy 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151
|||||
Db 1293 GTTGATCTTGGCTATGCTTACTTAAAGGCAAAAAAGTTCACTTTAAAGAAAGTAAACA 1352
|||||
Qy 152 IleLysGly-----LeuLeuLeuValGluAlaAspTyrThrThrLysAlaThr 167
|||||
Db 1353 ATAGGTGACAAACGTACATTGACATTGCAATCAACTGCAAAATATATCTTCTCAAGCACAC 1412
|||||
Qy 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179
|||||
Db 1413 GCAATCTTTACGGTTTGAATTAATATAGTTTC 1448
|||||
RESULT 7
US-08-472-172-11
; Sequence 11, Application US/08472172
; Patent No. 5985288
; GENERAL INFORMATION:
; APPLICANT: Munson, Jr., Robert S
; APPLICANT: Grass, Susan
; APPLICANT: Chong, Pele-Y Y
; APPLICANT: Fahim, Raafat
; APPLICANT: Sia, Charles D Y
; APPLICANT: McVerry, Patrick
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Outer Membrane Protein p1 and Peptides
; TITLE OF INVENTION: of Haemophilus influenzae Type B
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R6
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,172
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/849,411
; FILING DATE: 07-JUL-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-471
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; TELEX: 063-24567 SIMBAS
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 base pairs
; TYPE: nucleic acid
```



Db	1357	TCATTGGCGCTT	AGGTGGCGGGTATATGTTACACCTGAATGTCGCTTTTGGGT---	1411
Qy	137	AlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsnLeuGlyLeuLeu	156	
Db	1414	-----GTTAAATAC	TTTAAATTTGGAGATGCTCAA-----	1443
Qy	157	LeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeuTyrGlyLeuAsnLeuAsn	176	
Db	1444	-----GCACAGCTACCAACCAAGATAAAGTAGGTAACTTTTGATAGTAATGATGGC	1494	
Qy	177	Tyr	177	
Db	1495	TAT	1497	
RESULT 10				
PCT-US95-05134-11				
; Sequence 11, Application PC/TUS9505134				
; GENERAL INFORMATION:				
; APPLICANT: Murphy, Timothy F.				
; APPLICANT: Bhushan, Reva				
; TITLE OF INVENTION: Vaccine for Moraxella catarrhalis				
; NUMBER OF SEQUENCES: 18				
; CORRESPONDENCE ADDRESS:				
; ADDRESSEE: Hodgson, Russ, Andrews, Woods & Goodyear				
; STREET: 1800 One M&T Plaza				
; CITY: Buffalo				
; STATE: New York				
; COUNTRY: United States				
; ZIP: 14203-2391				
; COMPUTER READABLE FORM:				
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Kb storage				
; COMPUTER: IBM Compatible				
; OPERATING SYSTEM: MS-DOS/ Microsoft Windows 3.1				
; SOFTWARE: Wordperfect				
; CURRENT APPLICATION DATA:				
; FILING DATE: PCT/US95/05134				
; PRIOR APPLICATION DATA:				
; APPLICATION NUMBER: U.S. Serial No. 08/245,758				
; FILING DATE: 17/05/94				
; ATTORNEY/AGENT INFORMATION:				
; NAME: Neilson, M. Bud				
; REGISTRATION NUMBER: 35,300				
; REFERENCE/DOCKET NUMBER: 11520.0063				
; TELECOMMUNICATION INFORMATION:				
; TELEPHONE: (716) 856-4000				
; TELEFAX: (716) 849-0349				
; INFORMATION FOR SEQ ID NO: 11:				
; SEQUENCE CHARACTERISTICS:				
; LENGTH: 1650 nucleotides				
; TYPE: nucleic acid				
; STRANDEDNESS: double-stranded				
; TOPOLOGY: linear				
; MOLECULE TYPE: genomic DNA				
; HYPOTHETICAL: yes				
; IMMEDIATE SOURCE:				
; LIBRARY: genomic				
; CLONE: EMBL-3 clone				
; ORIGINAL SOURCE:				
; ORGANISM: Moraxella catarrhalis				
; STRAIN: 25240				
; CELL TYPE: bacterium				
; FEATURE:				
; LOCATION: E gene region, 154-1531				
PCT-US95-05134-11				
Alignment Scores:				
Pred. No.: 0.0638				
Score: 89.00				
Length: 1650				
Matches: 42				
Percent Similarity: 33.70%				
Conservative: 19				
Best Local Similarity: 23.20%				
Mismatch: 84				
Indels: 36				
Query Match: 9.30%				
DB: 5				
US-09-995-493-52 (1-179) x US-08-810-655A-1 (1-1650)				
Qy	9	TyrIleGlyLysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPhe	28	
Db	1027	TATCGCGGTAAAGAGGATTTTAAAGTTACTTTGGCTGACTGACTGAACTTATTTTCAA	1086	
Qy	29	HisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArg	48	
Db	1087	ACTGGTGTGATCAACACACCGCTATTACTCCAAAGTACCGTATGTACCATGCTCTGAT	1146	
Qy	49	PheLysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLys	68	
Db	1147	TTTCACATTCGCCCAACACAGATATACAGAA-----ACCACAAA	1185	
Qy	69	GluGluTyr-----LysAspAsnSerArgPheala	78	
Db	1186	CTTCGTTATCACAGGGTTTACCANTCATCATGATGACAAAGACCAATGTCGGCTGAA	1245	
Qy	79	IleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAsp	98	
Db	1246	GTTCGTTGGGTAAAGCGTTTAGCGATCGTTTGGCTGTTTCCAGGTGCGGTAGGTGGGAT	1305	
Qy	99	LysAlaAlaSerLysThrHisLeuSerAlaSerIleProAspThrAspArgMetTrpTyr	118	
Db	1306	AGTGGTCGAGGTAC-----CCTGCAAGTAGCTTAGTTCCTATCAAGCGCTATTAT	1356	
Qy	119	Ser-----IleGlyAlaThrTyrLysPheThrProAsnLeuSerValAspValcIlePhe	136	

US-09-995-493-52 (1-179) x PCT-US95-05134-11 (1-1650)

QY 9 TyrIleGlySerGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPhe 28  
DB 1027 TATCGGGTAAAGGATTTAAAGTTTACCTTGGCTGACATGCACTAGATTTTCAA 1086  
QY 29 HisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArg 48  
DB 1087 ACTGGTGTGAATCCAAACGCTATTAACTGCCAAAGTACGCTATGATACCTGCTGAT 1146  
QY 49 PheLysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLys 68  
DB 1147 TTTGACATTCGCCCAACACAGTATACAGAA-----ACCACAAA 1185  
QY 69 GluGluTyr-----LysAspAsnSerArgPheAla 78  
DB 1186 CTTCGTTATCCACAGGTTTACCATCATCAGCTATGACAAAGACCAATGTCGGTGAA 1245  
QY 79 IleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAsp 98  
DB 1246 GTTGGTTGGTAAGCGTGTAGCGATCGTTTGGCTGTTTCAGGTGCGGTAGGTGGAT 1305  
QY 99 LysAlaLeuSerLysThrHisLeuSerAlaSerIleProAspThrAspArgMetTrpTyr 118  
DB 1306 AGTGGTCGAGGTAAAC-----CCTGCAAGTAGCTTAGGTCCTATCAAGGCTATAT 1356  
QY 119 Ser-----IleGlyAlaThrTyrLysPheThrProAsnLeuSerValaspValGlyPhe 136  
DB 1357 TCATGGGCTTAGTGGCGCGGTATATGTTACACCTCAATGGTGGCTCTTGGGT--- 1413  
QY 137 AlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsnIleLysGlyLeuLeu 156  
DB 1414 -----GCTAAATACTTTAAATTTGGAGATGCTCAA----- 1443  
QY 157 LeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeuTyrGlyLeuAsnLeuAsn 176  
DB 1444 -----GCACAGCTACCACCAACCAAGATAAGTAGGTAACTTTGATAGTAATGATGGC 1494  
QY 177 Tyr 177  
DB 1495 TAT 1497

RESULT 11  
US-09-453-702B-173/c  
Sequence 173, Application US/09453702B  
Patent No. 6365723  
GENERAL INFORMATION:  
APPLICANT: Blattner, Frederick R.  
Perna, Nicole T.  
Plunkett, Guy  
Welch, Rod  
TITLE OF INVENTION: No. 6365723el Sequences of E. coli 0157  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Plinkney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53701-2113  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 8.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/453,702B  
FILING DATE: 03-Dec-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/110,955

FILING DATE: 04-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296.95017  
TELEPHONE: (608) 251-5000  
TELEFAX: (608) 251-9166  
INFORMATION FOR SEQ ID NO: 173:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23451  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 173:  
US-09-453-702B-173

Alignment Scores:  
Pred. No.: 5.06 Length: 23451  
Score: 87.00 Matches: 49  
Percent Similarity: 36.67% Conservative: 28  
Best Local Similarity: 23.33% Mismatches: 63  
Query Match: 9.09% Indels: 70  
DB: 4 Gaps: 12  
US-09-995-493-52 (1-179) x US-09-453-702B-173 (1-23451)  
QY 9 TyrIleGlyLysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPhe 28  
DB 6332 TATGTTATGAGTGGTTCA-----AGACTGGCTGATAAT 6300  
QY 29 HisGlnLeuThrAspGlnTrpAla----- 36  
DB 6299 CATACGCTATCGGCTGCTTATGCTCAGAGTAAGGTACAGGACTTAAAGATATTAAAGGA 6240  
QY 37 IleHisTyrSerTyrLysTyrThrGluTrp-----SerArgPhe 49  
DB 6239 GTCAATTCGAATATTCGATAT-----GAGTGGGATTTCTCCGGTACGTGTCGGGAGTTT 6183  
QY 50 LysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAla-----PheThr 66  
DB 6182 TCGTATATGAAGGTGATTGGCGGACAGTCATCGCCTGACGATTTTACAGA 6123  
QY 67 LysLysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThrTyrSerLeuAsn 86  
DB 6122 CATCAGGCTGATATAAATATTATTC-----TTTCTCGCAGGCCAGCTTATCGTCTCAAT 6066  
QY 87 AspAlaLeuThrLeu-----ArgAlaGlyLeuAlaTyrAspLysAla----- 100  
DB 6065 GACTATATCAGTTTTTATGTTGGTGGGATATCTCATATAAAGCGAAGGGAGATTAT 6006  
QY 101 -----AlaSerLysThrHisLeuSerAlaSerIle--- 110  
DB 6005 GAGTGGAGAACACTGTGGGTGCTGATGATGATATTTAAGTACTGCTGTATCA 5946  
QY 111 ---ProAspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsn 129  
DB 5945 AAAAAATCCACTGAT-----TTTGGCTTATGACAGTGGTCTTATTATTAATCCGTGGGCAAT 5889  
QY 130 LeuSerValaspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThr 149  
DB 5888 ATGCTCTGTAATCTGGGTAT-----GAGGGAATAA----- 5856  
QY 150 GlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsn 169  
DB 5855 -----GCAGATATATATGGAACATTCGCTCAAT 5826  
QY 170 LeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
DB 5825 GGATTTTACTGTTGGAGTTGGTTATCGTTTT 5796  
RESULT 12

Db	311	CTAGCTAAGACGGAATTAATGAGAAACTACTCTTGGTGGTGAAGACGTGCGGGGTGG	370
Qy	111	ProAspThrAspArgMetTtptYrSerIleGlyAlaThrTyrTyrLysPheThrProAsnLeu	130
Db	371	ATAACGATGATTCGACTTGGTACTACCTAGATCCAAACAACCTGGTATT-----	418
Qy	131	ServValAspValGlyPheAlaHisLeuArgGlyLysLysHisPheValGluThrGln	150
Db	419	--ATGCAACAGCTGGCAATATCTA--GGTAATAAGTGGTACTACTCGCTTCCTCA	472
Qy	151	AsnIleLysGlyLeuLeuLeuValGluAlaAspTyr	162
Db	473	GGAGC-AATGCCCACTGGCTGGTATCAGGAAGGTAC	507
RESULT 13			
US-08-183-214-1			
; Sequence 1, Application US/08183214			
; Patent No. 5716816			
GENERAL INFORMATION:			
; APPLICANT: Moss, Joel			
; APPLICANT: Stanley, Sally J.			
; APPLICANT: Nightingale, Maria S.			
; APPLICANT: Murtagh, Jr., James J.			
; APPLICANT: Monaco, Lucia			
; APPLICANT: Takada, Tatsuyuki			
; TITLE OF INVENTION: CLONES ENCODING MAMMALIAN			
; TITLE OF INVENTION: ADP-RIBOSYLARGININE HYDROLASES			
; NUMBER OF SEQUENCES: 12			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Townsend and Townsend			
; STREET: One Market Plaza, Steuart Street Tower			
; CITY: San Francisco			
; STATE: CA			
; COUNTRY: USA			
; ZIP: 94105			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: Patent In Release #1.0, Version #1.25			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/183,214			
; FILING DATE: 14-JAN-1994			
; CLASSIFICATION: 435			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 07/888,231			
; FILING DATE: 22-MAY-1992			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Weber, Kenneth A.			
; REGISTRATION NUMBER: 31,677			
; REFERENCE/DOCKET NUMBER: 15280-27			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: 415-543-9600			
; TELEFAX: 415-543-5043			
; INFORMATION FOR SEQ ID NO: 1:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 1245 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: single			
; TOPOLOGY: linear			
; MOLECULE TYPE: cDNA			
; FEATURE:			
; NAME/KEY: misc_signal			
; LOCATION: 1..25			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: 27..1115			
US-08-183-214-1			
Alignment Scores:			
Pred. No.:	0.22	Length:	1245
Score:	83.50	Matches:	36
Percent Similarity:	40.74%	Conservative:	19





ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII text  
CURRENT APPLICATION DATA:  
FILING DATE: US/08/961.527  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
FILING DATE:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 146:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11887 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-146

Alignment Scores:  
Pred. No.: 5.45 Length: 11887  
Score: 83.50 Matches: 42  
Percent Similarity: 39.53% Conservative: 26  
Best Local Similarity: 24.42% Mismatches: 77  
Query Match: 8.73% Indels: 28  
DB: 4 Gaps: 7

US-09-995-493-52 (1-179) x US-08-961-527-146 (1-11887)

Qy 8 ProTyrLeuGlyGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGly 27  
Db 11558 CGGTTTCCATCTAAAGGTAGTACAAATGGCTTACCACCAATGGTATCAGATTAGAAGT 11499  
Qy 28 PheHisGlnLeuThrAspGlnTrpAlaLeuHisTyrSerTyrLys-----TyrThr 44  
Db 11498 TTTCACAAAG-----TCAGAGTGG-----TACTACTTCGATAAAATCGAGTGTACAA 11451  
Qy 45 GluTrpSerArgPheLysLeuArgGlyLysTyrGlnAspGlySerClyTyrGluAla 64  
Db 11450 GAGTTTGGTTGGAAACATTAGAGATTAAACTAAGACAGGTGGTGGGAAGAAGTAC 11391  
Qy 65 PheThrLysLysGluGluTyrLysAspAsnSerArgPheAlaLeuGlyThrTyrSer 84  
Db 11390 GGGGAAACGTTGAAGATTCAAGACATAAAGAAGAGACCGTTATTATACGNACTATTAC 11331  
Qy 85 LeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAla----- 100  
Db 11330 TTTAATCAAAATCATCTCTTTAGACAGGTTGGCTTTTATGATCACTCAACTGGTATTAT 11271  
Qy 101 AlaSerLysThrHisLeuSer-----AlaSerIle 110  
Db 11270 CTAGCTAGACGGAATTAATTAAGGAAACACTACCTGGTGGTGAAGACGTCGGGGTGG 11211  
Qy 111 ProAspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeu 130  
Db 11210 ATAAACGATGATTCGACTTGGTACTACCTAGATCAACAACACTGGTATT----- 11163  
Qy 131 SerValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGln 150  
Db 11162 ---ATGCAACAGGTTGGCAATATCTA---GGTAATAAGTGGTACTACCTCGTTCCTCA 11109  
Qy 151 AsnIleLysGlyLeuLeuValGluAlaAspTyr 162  
Db 11108 GGAGC-AATGGCCACTGGCTGGTATACGGAAGGTAC 11074

## RESULT 16

US-09-453-702B-194/G  
Sequence 194, Application US/09453702B  
Patent No. 6365723  
GENERAL INFORMATION:  
APPLICANT: Blattner, Frederick R.  
Burland, Valerie  
Perna, Nicole T.  
Plunkett, Guy  
Weich, Rod  
TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Pinckney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53701-2113  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 8.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/453,702B  
FILING DATE: 03-Dec-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/110,955  
FILING DATE: 04-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296.95017  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 251-5000  
TELEFAX: (608) 251-9166  
INFORMATION FOR SEQ ID NO: 194:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9057  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 194:  
US-09-453-702B-194

Alignment Scores:  
Pred. No.: 4.29 Length: 9057  
Score: 83.00 Matches: 42  
Percent Similarity: 37.02% Conservative: 25  
Best Local Similarity: 23.20% Mismatches: 58  
Query Match: 8.67% Indels: 56  
DB: 4 Gaps: 10

US-09-995-493-52 (1-179) x US-09-453-702B-194 (1-9057)

Qy 33 AspGlnTrpAlaIleHis-----TyrSerTyrLysTyrThrGluTrpSerArgPheLys 50  
Db 1897 CACAAATGGTCATCTCGTGGGGGATGACTATCAATCCGACTCAACTGGCTGATG----- 1844  
Qy 51 GluLeuArgGlyLysTyrClnAspGlySerGlyTyrGluAlaPheThrLysGluGlu 70  
Db 1843 ---TTATTTGGCTCATAT-----GCCACGAGCATTCGCGCCCGACGATGGCGAAATG 1793  
Qy 71 TyrLysAspAsnSerArgPheAlaIleGlyThrTyrSer----- 84  
Db 1792 TATACGATTCTAACCACTTCTCGATTGGTCGCTTCTATACCACTATTGGGTGCCAATC 1733  
Qy 84 ----- 84

```
Db 1732 CCGAAGTCTAGCTCCGGAAGAACTACAGAACTACAGGATTCGGTTCGCTTTTGAT 1673
QY 85 -----LeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAla 100
Db 1672 GACCTGATGTTGCTCAATGATGCTCTGGAATTTAAAGCCAGCTACTTTGATACCAAGCG 1613
QY 101 -----AlaSerLysThrHisLeuSerAlaSerIle 110
Db 1612 AAGGATTACATCTCCAGACCGTCGATTTCGCGCGCGGACGACTATGCTGATAACGTC 1553
QY 111 proAspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeu 130
Db 1552 CCGAAGCCG---AAAATCTGG---GGCTGGGATGTGATGACGAAATATACCACTGATCTG 1499
QY 131 ---SerValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThr 149
Db 1498 TTTAGCTTGATGCTGCGCTATACCTACCGCGGCGGAAA-----GACACC 1454
QY 150 GlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsn 169
Db 1453 GATACCGCGGAATACATCTCCAGCATTAACCCGGAT---ACTGTTACGACACTCTGAAT 1397
QY 170 Leu 170
Db 1396 ATT 1394

RESULT 17
US-08-183-214-9
; Sequence 9, Application US/08183214
; Patent No. 5716816
; GENERAL INFORMATION:
; APPLICANT: Moss, Joel
; APPLICANT: Stanley, Sally J.
; APPLICANT: Nightingale, Maria S.
; APPLICANT: Murtash, Jr., James J.
; APPLICANT: Monaco, Lucia
; APPLICANT: Takada, Tatsuyuki
; TITLE OF INVENTION: CLONES ENCODING MAMMALIAN
; TITLE OF INVENTION: ADP-RIBOSYLARGININE HYDROLASES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend
; STREET: One Market Plaza, Stewart Street Tower
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/183,214
; FILING DATE: 14-JAN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/888,231
; FILING DATE: 22-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-27
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
```

```
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 10..1095
US-08-183-214-9
Alignment Scores:
Pred. No.: 0.335 Length: 1103
Score: 81.50 Matches: 37
Percent Similarity: 40.44% Conservative: 18
Best Local Similarity: 27.21% Mismatches: 54
Query Match: 8.52% Indels: 27
DB: 1 Gaps: 8

US-09-995-493-52 (1-179) x US-08-183-214-9 (1-1103)
QY 11 GLYLGLYSerLeuThrLeuLysLeuPro-----AlaTyrTrpGluLeuSerGly 27
Db 598 GGAAGAGGGCTAATGAGGTG---CTGCCTGAAGCCAAAGTACATCACTCAGTCAGGC 654
QY 28 PheHisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSer 47
Db 655 TAC-----TTTGTGAAGGAGAATCTTCAACACTGGTCTCTTCCGAGAAAGATGGAA 708
QY 48 ArgPheLysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLys 67
Db 709 AAGTACCTGGAGCTTAGAGGA---ATTTTGGACGGCAACTCGGCTCCCGCTTTCCCGCAG 765
QY 68 LysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThrThrTyrSerJ----- 84
Db 766 CCCTTTGGTGTGAAGGAAGGGATCAGTTCTACATCGACGTGAGCTACTCGGCTGGGT 825
QY 85 -----LeuAsnAspAlaLeuThrLeuArgAlaGly 94
Db 826 GGCAGCAGCGGACACGATGCCCCCATGATTGCCCTACGATGCCCTC---CTGGCCGCGGG 882
QY 95 LeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIleProAspThrAsp 114
Db 883 GATTCCTGGAAGGAGCTCGCACAGAGCGCTCTTCCACGGTGGAGACAGTGTATCCACG 942
QY 115 Arg-----MetTrpTyrSerIleGlyAlaThrTyrLysPhe 126
Db 943 GCGGCCATGCTCGCTCGCTGGTGG-----GGAGTTATGTACGGCTTT 984

RESULT 18
US-08-286-767-1
; Sequence 1, Application US/08286767
; Patent No. 5733760
; GENERAL INFORMATION:
; APPLICANT: Lu, Yichen
; APPLICANT: Miller, Samuel I.
; APPLICANT: Killen, Kevin
; TITLE OF INVENTION: SALMONELLA VECTORS ENCODING TRUNCATED
; TITLE OF INVENTION: PagC FUSION PROTEIN, METHOD OF MAKING, AND USES THEREOF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RONALD I. EISENSTEIN; DIKE, BRONSTEIN,
; ADDRESSEE: ROBERTS & CUSHMAN
; STREET: 130 Water Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,767
; FILING DATE: 05-AUG-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Resnick, David S.
```

```
; REGISTRATION NUMBER: 34235
; REFERENCE/DOCKET NUMBER: 44420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 523-3400
; TELEFAX: (617) 523-6400
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2319 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 729..1292
US-08-286-767-1

Alignment Scores:
Pred. No.: 1.12 Length: 2319
Score: 81.00 Matches: 49
Percent Similarity: 37.56% Conservative: 28
Best Local Similarity: 23.90% Mismatches: 60
Query Match: 8.46% Indels: 69
DB: 11 Gaps: 11

US-09-995-493-52 (1-179) x US-08-286-767-1 (1-2319)
Qy 17 LeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAla 36
Db 757 TTATTACTACAGCGTTTGGTT-GTAAATGTTGCACAGCGCGATACCTAACGCCCTTTTCC 815
Qy 37 IleHisTyrSer---TyrLysTyrThrGluTrpSerArgPheLysGluLeuArgGly--- 54
Db 816 GTGGGTATGCACGGTATGCACAAAGTAAAGTTTCAGGATTTCAAAAATATCCGAGGGTA 875
Qy 55 -----LysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGlu 70
Db 876 AATGTGAATACCGTTATGAGGATGACTCTCCGGTA---AGTTTATTTCCTCGCTAAGT 932
Qy 71 TyrLysAspAsnSerArgPheAla----- 78
Db 933 TACTTATATGGAGACAGACAGCGCTCCGGTCTGTTCAGCCCTGAAGGTATTCATTACCAT 992
Qy 79 -----IleGlyThrThrTyrSerLeuAsnAsp 87
Db 993 GACAACTTTGAGGTGAAGTACGGTTCTTTAATGTTGGCCAGCCCTATCGATTGCTGAC 1052
Qy 88 AlaLeuThrLeu-----ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHis 105
Db 1053 AATTTTTCGTTATACCGCGTGCAGCGGTGTCGCGACGGTAAGGCGCATTTAAAGAACAT 1112
Qy 106 LeuSerAlaSerIleProAspThrAsp-----ArgMet 116
Db 1113 -----TCCACTCAGATGGCGATCTTTTCTAACAAAAATTCCTCAGGAAAAACG 1163
Qy 117 TrpTyrSerIleGlyAlaThrTyrLysPheThrPro-----AsnLeuSerValAspVal 134
Db 1164 GGATTGCGTGGCGCGGTGTACAGATGAATCCGCTCGAGAATATCGTCGTCGATGTT 1223
Qy 135 GlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsnIleLysGly 154
Db 1224 GGGTAT-----GAAGGAAGCAACATCTCC--- 1247
Qy 155 LeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeuTyrGlyLeuAsn 174
Db 1248 -----TCTACAAAA-----ATAACGGCTTCAACGTCGGG 1277
Qy 175 LeuAsnTyrArgPhe 179.
Db 1278 GTTGATACCGTTTC 1292

RESULT 19
US-08-090-526-1

; Sequence 1, Application US/08090526
; Patent No. 5599537
; GENERAL INFORMATION:
; APPLICANT: Miller, Samuel I.
; TITLE OF INVENTION: SALMONELLA VIRULENCE GENES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/090,526
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/629,602
; FILING DATE: 12/18/90
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/192001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2320
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-090-526-1

Alignment Scores:
Pred. No.: 1.12 Length: 2320
Score: 81.00 Matches: 49
Percent Similarity: 37.56% Conservative: 28
Best Local Similarity: 23.90% Mismatches: 60
Query Match: 8.46% Indels: 69
DB: 11 Gaps: 11

US-09-995-493-52 (1-179) x US-08-090-526-1 (1-2320)
Qy 17 LeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAla 36
Db 757 TTATTACTACAGCGTTTGGTT-GTAAATGTTGCACAGCGCGATACCTAACGCCCTTTTCC 815
Qy 37 IleHisTyrSer---TyrLysTyrThrGluTrpSerArgPheLysGluLeuArgGly--- 54
Db 816 GTGGGTATGCACGGTATGCACAAAGTAAAGTTTCAGGATTTCAAAAATATCCGAGGGTA 875
Qy 55 -----LysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGlu 70
Db 876 AATGTGAATACCGTTATGAGGATGACTCTCCGGTA---AGTTTATTTCCTCGCTAAGT 932
Qy 71 TyrLysAspAsnSerArgPheAla----- 78
Db 933 TACTTATATGGAGACAGACAGCGCTCCGGTCTGTTCAGCCCTGAAGGTATTCATTACCAT 992
Qy 79 -----IleGlyThrThrTyrSerLeuAsnAsp 87
Db 993 GACAACTTTGAGGTGAAGTACGGTTCTTTAATGTTGGCCAGCCCTATCGATTGCTGAC 1052
Qy 88 AlaLeuThrLeu-----ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHis 105
Db 1053 AATTTTTCGTTATACCGCGTGCAGCGGTGTCGCGACGGTAAGGCGCATTTAAAGAACAT 1112
Qy 106 LeuSerAlaSerIleProAspThrAsp-----ArgMet 116
Db 1113 -----TCCACTCAGATGGCGATCTTTTCTAACAAAAATTCCTCAGGAAAAACG 1163
Qy 117 TrpTyrSerIleGlyAlaThrTyrLysPheThrPro-----AsnLeuSerValAspVal 134
Db 1164 GGATTGCGTGGCGCGGTGTACAGATGAATCCGCTCGAGAATATCGTCGTCGATGTT 1223
Qy 135 GlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsnIleLysGly 154
Db 1224 GGGTAT-----GAAGGAAGCAACATCTCC--- 1247
Qy 155 LeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeuTyrGlyLeuAsn 174
Db 1248 -----TCTACAAAA-----ATAACGGCTTCAACGTCGGG 1277
Qy 175 LeuAsnTyrArgPhe 179.
Db 1278 GTTGATACCGTTTC 1292

RESULT 19
US-08-090-526-1
```

Db 1053 AATTTTCTGTTATACGGCTGGCGGTGTCTGGCACGGTAAGGCGACATTAAAGAACAT 1112  
Qy 106 LeuSerAlaSerIleProAspThrAsp-----ArgMet 116  
Db 1113 -----TCCACTCAGGATGGCGATCTTTTCTAACAAAATTTCTCCTCAAGGAACG 1163  
Qy 117 TrpTyrSerIleGlyAlaThrTyrLysPheThrPro-----AsnLeuSerValAspVal 134  
Db 1164 GGATTTCCTGGCGCGGGGTGTACACATGAATCCGCTGGAGAAATATCGTCGTGATGTT 1223  
Qy 135 GlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsnIleLysGly 154  
Db 1224 GGGTAT-----GAAGGAACCAACATCTCC--- 1247  
Qy 155 LeuLeuValGluAlaAspTyrThrLysAlaThrLysAlaAsnLeuTyrGlyLeuAsn 174  
Db 1248 -----TCTACAAA-----ATAACGGCTTCACGTCGGG 1277  
Qy 175 LeuAsnTyrArgPhe 179  
Db 1278 GTTGGATACCGTTTC 1292

## RESULT 20

US-08-486-715-1  
; Sequence 1, Application US/08486715  
; Patent No. 5674724  
; GENERAL INFORMATION:  
; APPLICANT: Miller, Samuel I.  
; TITLE OF INVENTION: SALMONELLA VIRULENCE GENES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P. C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 502 or 55SX  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/486,715  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/090,526  
; FILING DATE: 09-JUL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/629,602  
; FILING DATE: 18-DEC-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00786/192005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2320  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-486-715-1  
Alignment Scores:  
Pred. No.: 1.12 Length: 2320  
Score: 81.00 Matches: 49  
Percent Similarity: 37.56% Conservativeness: 28  
Best Local Similarity: 23.90% Mismatches: 60

Query Match: 8.46% Indels: 69  
DB: 1 Gaps: 11  
US-09-995-493-52 (1-179) x US-08-486-715-1 (1-2320)  
Qy 17 LeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAla 36  
Db 757 TTATTACTACAAGCGTTTGGTT-GTAAATGTTGCACAGCGCGATACGACGCGCTTTTCC 815  
Qy 37 IleHisTyrSer---TyrLysTyrThrGluTrpSerArgPheLysGluLeuArgGly--- 54  
Db 816 GTGGGGTATCACGGTATGCACAAAGTAAAGTTCAGGATTTCAAAATAATATCCGAGGGTA 875  
Qy 55 -----LysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGlu 70  
Db 876 AATCTGAATACCGTTATGAGGATGACTCTCCGCTA---AGTTTATTCTCCTCCCTAAGT 932  
Qy 71 TyrLysAspAsnSerArgPheAla----- 78  
Db 933 TACTTATATGACACACAGCGCTTCGGGTCTGTTGAGCTGAAGGTATTCATTACCAT 992  
Qy 79 -----IleGlyThrThrTyrSerLeuAsnAsp 87  
Db 993 GACAAGTTTGAAGTGAAGTACGGTCTTTAATGTTGGCGCAGCTATCGATTCTCTGAC 1052  
Qy 88 AlaLeuThrLeu-----ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHis 105  
Db 1053 AATTTTTCGTTATACGGCTGGCGGTGTCGGCAGCGTAAGCGGACATTTAAAGAACAT 1112  
Qy 106 LeuSerAlaSerIleProAspThrAsp-----ArgMet 116  
Db 1113 -----TCCACTCAGGATGGCGATTCTTTTCTAACAAAATTTCTCCTCAAGGAACG 1163  
Qy 117 TrpTyrSerIleGlyAlaThrTyrLysPheThrPro-----AsnLeuSerValAspVal 134  
Db 1164 GGATTTCCTGGCGCGGGGTGTACACATGAATCCGCTGGAGAAATATCGTCGTGATGTT 1223  
Qy 135 GlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsnIleLysGly 154  
Db 1224 GGGTAT-----GAAGGAACCAACATCTCC--- 1247  
Qy 155 LeuLeuValGluAlaAspTyrThrLysAlaThrLysAlaAsnLeuTyrGlyLeuAsn 174  
Db 1248 -----TCTACAAA-----ATAACGGCTTCACGTCGGG 1277  
Qy 175 LeuAsnTyrArgPhe 179  
Db 1278 GTTGGATACCGTTTC 1292

## RESULT 21

US-08-486-719-1  
; Sequence 1, Application US/08486719  
; Patent No. 5674736  
; GENERAL INFORMATION:  
; APPLICANT: Miller, Samuel I.  
; TITLE OF INVENTION: SALMONELLA VIRULENCE GENES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P. C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 502 or 55SX  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/486,719  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435



```
Qy 71 TyrLysAspAsnSerArgPheAla-----1247
Db 933 TACTTATATGAGACAGACAGGCTTCGGGGTCTCTTGGAGCTGAAGGTATTCATTACCAT 992
Qy 79 -----1247
Db 993 GACAAGTTTGAAGTGAAGTCTCTTAAATGTTGGCCAGCCTATCGATTCTCTGAC 1052
Qy 88 AlaLeuThrLeu-----ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHis 105
Db 1053 AATTTCGTATACGGCTGGCGGGTGTCTGGCAGCGTAAGAGGCGACATTTAAAGACAT 1112
Qy 106 LeuSerAlaSerIleProAspThrAsp-----ArgMet 116
Db 1113 -----TCCACTCAGGATGGCGATTCTTTTCTAACAAAATTTCTCCTCAAGGAAACG 1163
Qy 117 TrpTyrSerIleGlyAlaThrTyrLysPheThrPro-----AsnLeuSerValAspVal 134
Db 1164 GGATTTGCTGGCGGGGGGTGTACAGATGATCCGCTGGAGAAATATCGCTCGATGTT 1223
Qy 135 GlyPheAlaHisLeuArgGlyLysLysHisPheValGluThrGlnAsnIleLysGly 154
Db 1224 GGGTAT-----GAAGGAGCAACATCTCC---1247
Qy 155 LeuLeuValGluAlaAspTyrThrLysAlaThrLysAlaAsnLeuTyrGlyLeuAsn 174
Db 1248 -----TCTACAAA-----ATAAACGGCTTCAACGTCGGG 1277
Qy 175 LeuAsnTyrArgPhe 179
Db 1278 GTTGGATACCGTTTC 1292

RESULT 24
US-08-476-100-1
: Sequence 1, Application US/08565861
: Patent No. 5843426
: GENERAL INFORMATION:
: APPLICANT: Miller, Samuel I.
: APPLICANT: Mekalanos, John J.
: APPLICANT: Hohmann, Elizabeth
: TITLE OF INVENTION: SALMONELLA VACCINES
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street
```

```
LENGTH: 2320
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-476-100-1

Alignment Scores:
Pred. No.: 1.12 Length: 2320
Score: 81.00 Matches: 49
Percent Similarity: 37.56% Conservative: 28
Best Local Similarity: 23.90% Mismatches: 60
Query Match: 8.46% Indels: 69
DB: 1 Gaps: 11

US-09-995-493-52 (1-179) x US-08-476-100-1 (1-2320)
Qy 17 LeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAla 36
Db 757 TTATTAATAAAGCGTTTGGTT-GTAAATGTTGCACAGCCGATCTAAGCCCTTTTCC 815
Qy 37 IleHisTyrSer---TyrLysTyrThrGluTrpSerArgPheLysGluLeuArgGly--- 54
Db 816 GTGGGGTATGCACGGTATGCACAAAGTAAAGTTCAGGATTTCAAAAATATCCGAGGGTA 875
Qy 55 -----LysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGlu 70
Db 876 AATGTGAATACCGTTATGAGGATGACTCTCCGGTA---AGTTTATTTCCTCGCTAAGT 932
Qy 71 TyrLysAspAsnSerArgPheAla-----78
Db 933 TACTTATATGAGACAGACAGGCTTCGGGGTCTCTTGGAGCTGAAGGTATTCATTACCAT 992
Qy 79 -----IleGlyThrThrTyrSerLeuAsnAsp 87
Db 993 GACAAGTTTGAAGTGAAGTACGGTTCTTTAATGTTGGCCAGCCTATCGATTCTCTGAC 1052
Qy 88 AlaLeuThrLeu-----ArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHis 105
Db 1053 AATTTCGTATACGGCTGGCGGGTGTCTGGCAGCGTAAGAGGCGACATTTAAAGACAT 1112
Qy 106 LeuSerAlaSerIleProAspThrAsp-----ArgMet 116
Db 1113 -----TCCACTCAGGATGGCGATTCTTTTCTAACAAAATTTCTCCTCAAGGAAACG 1163
Qy 117 TrpTyrSerIleGlyAlaThrTyrLysPheThrPro-----AsnLeuSerValAspVal 134
Db 1164 GGATTTGCTGGCGGGGGGTGTACAGATGATCCGCTGGAGAAATATCGCTCGATGTT 1223
Qy 135 GlyPheAlaHisLeuArgGlyLysLysHisPheValGluThrGlnAsnIleLysGly 154
Db 1224 GGGTAT-----GAAGGAGCAACATCTCC---1247
Qy 155 LeuLeuValGluAlaAspTyrThrLysAlaThrLysAlaAsnLeuTyrGlyLeuAsn 174
Db 1248 -----TCTACAAA-----ATAAACGGCTTCAACGTCGGG 1277
Qy 175 LeuAsnTyrArgPhe 179
Db 1278 GTTGGATACCGTTTC 1292

RESULT 24
US-08-565-861-1
: Sequence 1, Application US/08565861
: Patent No. 5843426
: GENERAL INFORMATION:
: APPLICANT: Miller, Samuel I.
: APPLICANT: Mekalanos, John J.
: APPLICANT: Hohmann, Elizabeth
: TITLE OF INVENTION: SALMONELLA VACCINES
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street
```



```
DB: 3 Gaps: 11
US-09-995-493-52 (1-179) x US-08-475-749-1 (1-2320)
QY 17 LeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAla 36
Db 757 TTATTACTACAAGCGTTTGGTT-GTAAATGTTGCACAGCCGATACACGCTTTTCC 815
QY 37 IleHisTyrSer---TyrLysTyrThrGluTrpSerArgPheLysGluLeuArgGly--- 54
Db 816 GTGGGGTATCGACGGTATGCACAAAGTAAAGTTCAGGATTTCAAAATATCCGAGGGTA 875
QY 55 -----LysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGlu 70
Db 876 AATGTGAATACCGTTATGAGTACTCTCCGTA---AGTTTATTCTCCTCGCTAAGT 932
QY 71 TyrLysAspAsnSerArgPheAla----- 78
Db 933 TACTTATATGAGACAGACAGCGCTTCGGGCTCTGTTGAGCCTGAAGGTATTTCATTACCAT 992
QY 79 -----IleGlyThrThrTyrSerLeuAsnAsp 87
Db 993 GACAAGTTTCAGGTGAGTACGTTCTTTAATGTTGGCGCCAGCCTATCGATTGCTGAC 1052
QY 88 AlaLeuThrLeu-----ArgAlaGlyLeuAlaTyrAspLysAlaLaSerLysThrHis 105
Db 1053 AATTTTTCGTTATACGCGCTGCGGGTGTCCGACGGTAAAGCGACATTTAAAGAACAT 1112
QY 106 LeuSerAlaSerIleProAspThrAsp----- 116
Db 1113 -----TCCACTCAGGATGCGGATCTTTTCTAACAAAAATTTCTCAAGGAAACG 1163
QY 117 TrpTyrSerIleGlyAlaThrTyrLysPheThrPro-----AsnLeuSerValAspVal 134
Db 1164 GGATTTCCCTGGCGCGGGGTGTACAGATCAATCCGCTGGAGATATCGTCGCTGATT 1223
QY 135 GlyPheAlaHisLeuArgGlyLysLysLysPheValGluThrGlnAsnIleLysGly 154
Db 1224 GGCTAT-----GAAGGAAGCAACATCTCC--- 1247
QY 155 LeuLeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeuTyrGlyLeuAsn 174
Db 1248 -----TCTACAAA-----ATAAAGCGCTTCAACGTCGGG 1277
QY 175 LeuAsnTyrArgPhe 179
Db 1278 GTTGATACCGTTTC 1292
```

```
RESULT 26
PCT-US94-07658-1
; Sequence 1, Application PC/TUS9407658
; GENERAL INFORMATION:
; APPLICANT: Miller, Samuel I.
; APPLICANT: Mekalanos, John J.
; TITLE OF INVENTION: SALMONELLA VACCINES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version 5.1
; SOFTWARE: #1.25 and WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07658
; FILING DATE:
; CLASSIFICATION:
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/090,526
; FILING DATE: July 9, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/220001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2320
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US94-07658-1
Alignment Scores:
Pred. No.: 1,12 Length: 2320
Score: 81,00 Matches: 49
Percent Similarity: 37,56% Conservative: 28
Best Local Similarity: 23,90% Mismatches: 60
Query Match: 8,46% Indels: 69
DB: 5 Gaps: 11
US-09-995-493-52 (1-179) x PCT-US94-07658-1 (1-2320)
```

```
QY 17 LeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAla 36
Db 757 TTATTACTACAAGCGTTTGGTT-GTAAATGTTGCACAGCCGATACACGCTTTTCC 815
QY 37 IleHisTyrSer---TyrLysTyrThrGluTrpSerArgPheLysGluLeuArgGly--- 54
Db 816 GTGGGGTATCGACGGTATGCACAAAGTAAAGTTCAGGATTTCAAAATATCCGAGGGTA 875
QY 55 -----LysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGlu 70
Db 876 AATGTGAATACCGTTATGAGTACTCTCCGTA---AGTTTATTCTCCTCGCTAAGT 932
QY 71 TyrLysAspAsnSerArgPheAla----- 78
Db 933 TACTTATATGAGACAGACAGCGCTTCGGGCTCTGTTGAGCCTGAAGGTATTTCATTACCAT 992
QY 79 -----IleGlyThrThrTyrSerLeuAsnAsp 87
Db 993 GACAAGTTTCAGGTGAGTACGTTCTTTAATGTTGGCGCCAGCCTATCGATTGCTGAC 1052
QY 88 AlaLeuThrLeu-----ArgAlaGlyLeuAlaTyrAspLysAlaLaSerLysThrHis 105
Db 1053 AATTTTTCGTTATACGCGCTGCGGGTGTCCGACGGTAAAGCGACATTTAAAGAACAT 1112
QY 106 LeuSerAlaSerIleProAspThrAsp----- 116
Db 1113 -----TCCACTCAGGATGCGGATCTTTTCTAACAAAAATTTCTCAAGGAAACG 1163
QY 117 TrpTyrSerIleGlyAlaThrTyrLysPheThrPro-----AsnLeuSerValAspVal 134
Db 1164 GGATTTCCCTGGCGCGGGGTGTACAGATCAATCCGCTGGAGATATCGTCGCTGATT 1223
QY 135 GlyPheAlaHisLeuArgGlyLysLysLysPheValGluThrGlnAsnIleLysGly 154
Db 1224 GGCTAT-----GAAGGAAGCAACATCTCC--- 1247
QY 155 LeuLeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeuTyrGlyLeuAsn 174
Db 1248 -----TCTACAAA-----ATAAAGCGCTTCAACGTCGGG 1277
QY 175 LeuAsnTyrArgPhe 179
Db 1278 GTTGATACCGTTTC 1292
```



## RESULT 27

US-09-838-169  
: Sequence 169, Application US/09370838  
: Patent No. 644425  
: GENERAL INFORMATION:  
: APPLICANT: Reed, Steven G.  
: APPLICANT: Lodes, Michael J.  
: APPLICANT: Mohamath, Roadoh  
: APPLICANT: Secrist, Heather  
: TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF  
: LUNG CANCER AND METHODS FOR THEIR USE  
: FILE REFERENCE: 210121.475C1  
: CURRENT APPLICATION NUMBER: US/09/370.838  
: CURRENT FILING DATE: 1999-08-09  
: EARLIER APPLICATION NUMBER: US 09/285,323  
: EARLIER FILING DATE: 1999-04-02  
: NUMBER OF SEQ ID NOS: 289  
: SOFTWARE: FastSeq for Windows Version 3.0  
: SEQ ID NO 169  
: LENGTH: 547  
: TYPE: DNA  
: ORGANISM: Homo sapien  
US-09-370-838-169

Alignment Scores:  
Pred. No.: 0.193 Length: 547  
Score: 80.00 Matches: 34  
Percent Similarity: 40.16% Conservative: 15  
Best Local Similarity: 27.87% Mismatches: 53  
Query Match: 8.36% Indels: 20  
DB: 4 Gaps: 6

US-09-995-493-52 (1-179) x US-09-370-838-169 (1-547)

Qy	32	ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu	51
Db	124	ACTGTGAGATGGTGGCAGTGTGGAGCATGAGGCCACTAAGTGCAGAGTTTCCGCGAC	183
Qy	52	---LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGlu	70
Db	184	CATATGAAAGCGTCATCCCATCCGATGTCCTGCTGTGTAAGAAAGCCCTCC	243
Qy	71	TyrLysAspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThr	90
Db	244	TACCTTGATTGTCATCAGG---GCCATTGGCGCAACGAAGCG-----GATGCTGTGACA	294
Qy	91	LeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSer-----	107
Db	295	CTGGATGCAGGTTTGGTGTATGATGCTTACCTGGCTCCCAATAACCTGAAGCCTGTGGTG	354
Qy	108	-----AlaSerIleProAspThrAspArgMetTrpTyrSerIleGlyAlaThr	123
Db	355	GCAGAGTTCTATGGGTCAAAAGAGGATCCACAGACTTTCTATATGCTTCTGTGGTG	414
Qy	124	TyrLysPheThrProAsnLeuSerValAspValGlyPheAla-----HisLeuArgGly	141
Db	415	AAGAAG-----GATAGTGGCTTCCAGATGAACCAAGCTTCGAGGC	453
Qy	142	LysLys 143	
Db	454	AAGAAG 459	

## RESULT 28

US-08-848-760B-25  
: Sequence 25, Application US/08848760B  
: Patent No. 6248721  
: GENERAL INFORMATION:  
: APPLICANT: Chang, Lung-Ji  
: TITLE OF INVENTION: Animal Model For Evaluation Of Vaccines  
: NUMBER OF SEQUENCES: 33  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: United States of America  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/848,760B  
APPLICATION NUMBER: 08/848,760B  
FILING DATE: 25-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/838,702  
FILING DATE: 09-APR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: PACE, DORAN R.  
REGISTRATION NUMBER: 38,261  
REFERENCE/DOCKET NUMBER: CNG-100C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (352) 375-8100  
TELEFAX: (352) 372-5800  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2097 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "DNA"  
SEQUENCE DESCRIPTION: SEQ ID NO: 25:  
US-08-848-760B-25

Alignment Scores:  
Pred. No.: 1.31 Length: 2097  
Score: 80.00 Matches: 34  
Percent Similarity: 40.16% Conservative: 15  
Best Local Similarity: 27.87% Mismatches: 53  
Query Match: 8.36% Indels: 20  
DB: 4 Gaps: 6

US-09-995-493-52 (1-179) x US-08-848-760B-25 (1-2097)

Qy	32	ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu	51
Db	70	ACTGTGAGATGGTGGCAGTGTGGAGCATGAGGCCACTAAGTGCAGAGTTTCCGCGAC	129
Qy	52	---LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGlu	70
Db	130	CATATGAAAGCGTCATCCCATCCGATGTCCTGCTGTGTAAGAAAGCCTCC	189
Qy	71	TyrLysAspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThr	90
Db	190	TACCTTGATTGTCATCAGG---GCCATTGGCGCAACGAAGCG-----GATGCTGTGACA	240
Qy	91	LeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSer-----	107
Db	241	CTGGATGCAGGTTTGGTGTATGATGCTTACTTGGCTCCCAATAACCTGAAGCCTGTGGTG	300
Qy	108	-----AlaSerIleProAspThrAspArgMetTrpTyrSerIleGlyAlaThr	123
Db	301	GCAGAGTTCTATGGGTCAAAAGAGGATCCACAGACTTTCTATATGCTGTGTGGTG	360
Qy	124	TyrLysPheThrProAsnLeuSerValAspValGlyPheAla-----HisLeuArgGly	141
Db	361	AAGAAG-----GATAGTGGCTTCCAGATGAACCAAGCTTCGAGGC	399
Qy	142	LysLys 143	
Db	400	AAGAAG 405	



Score: 80.00 Matches: 34  
Percent Similarity: 40.16% Conservative: 15  
Best Local Similarity: 27.87% Mismatches: 53  
Query Match: 8.36% Indels: 20  
Gaps: 6

US-09-995-493-52 (1-179) x US-09-037-188-1 (1-3427)

QY 52 ---LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGlu 70  
Db 1271 CATATGAAAAGCGTCATTCATCCGATGTCAGTGTTCGTTGCTGTAAGAAGCGCTCC 1330

QY 71 TyrLysAspAsnSerArgPheAlaIleGlyThrTyrSerLeuAsnAspAlaLeuThr 90  
Db 1331 TACCTTGATGTCATCAGG---GCCATTGCGCAAGCG-----GATGCTGTGACA 1381

QY 91 LeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSer----- 107  
Db 1382 CTGGATGCAGGTTGGTGATGATGCTTACTTGGCTCCCAATAACCTGAAGCGCTGGTG 1441

QY 108 -----AlaSerIleProAspThrAspArgMetTrpTyrSerIleGlyAlaThr 123  
Db 1442 GCAGAGTTCTATGGGTCAAAGAGGATCCACAGACTTTCTATTATGCTTGTGTTGGTG 1501

QY 124 TyrLysPheThrProAsnLeuSerValAspValGlyPheAla-----HisLeuArgGly 141  
Db 1502 AGAAG-----GATAGTGGCTTCCAGATGAACAGCTTCGAGGC 1540

QY 142 LysLys 143  
Db 1541 AGAAG 1546

RESULT 31  
US-09-037-188-1  
; Sequence 1, Application US/09037188  
; Patent No. 6027921  
; GENERAL INFORMATION:  
; APPLICANT: Heartlein, Michael W.  
; APPLICANT: Lemontt, Jeffrey F.  
; APPLICANT: Concino, Michael F.  
; TITLE OF INVENTION: CHIMERIC PROTEINS FOR USE  
; TITLE OF INVENTION: IN TRANSPORT OF A SELECTED SUBSTANCE INTO CELLS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; FILING DATE: 02-MAR-1998  
; FILING DATE: 02-MAR-1998  
; PRIOR APPLICATION NUMBER:  
; FILING DATE: 06-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fraser, Ph.D., J.D., Janis K.  
; REGISTRATION NUMBER: 34,819  
; REFERENCE/DOCKET NUMBER: 07236/009002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3427 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
; MOLECULE TYPE: cDNA to mRNA  
US-09-037-188-1

Alignment Scores: 2.63 Length: 3427  
Pred. No.:

Score: 80.00 Matches: 34  
Percent Similarity: 40.16% Conservative: 15  
Best Local Similarity: 27.87% Mismatches: 53  
Query Match: 8.36% Indels: 20  
Gaps: 6

US-09-995-493-52 (1-179) x US-09-037-188-1 (1-3427)

QY 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51  
Db 1211 ACTGTGAGATGGTGTCCAGTGTCCGAGCATGAGCCACTAAGTGCAGAGTTTCCGCGAC 1270

QY 52 ---LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGlu 70  
Db 1271 CATATGAAAAGCGTCATTCATCCGATGTCAGTGTTCGTTGTTGTGAAGAAGCGCTCC 1330

QY 71 TyrLysAspAsnSerArgPheAlaIleGlyThrTyrSerLeuAsnAspAlaLeuThr 90  
Db 1331 TACCTTGATGTCATCAGG---GCCATTGCGCAAGCG-----GATGCTGTGACA 1381

QY 91 LeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSer----- 107  
Db 1382 CTGGATGCAGGTTGGTGATGATGCTTACTTGGCTCCCAATAACCTGAAGCGCTGGTG 1441

QY 108 -----AlaSerIleProAspThrAspArgMetTrpTyrSerIleGlyAlaThr 123  
Db 1442 GCAGAGTTCTATGGGTCAAAGAGGATCCACAGACTTTCTATTATGCTGTTGTGGTG 1501

QY 124 TyrLysPheThrProAsnLeuSerValAspValGlyPheAla-----HisLeuArgGly 141  
Db 1502 AGAAG-----GATAGTGGCTTCCAGATGAACAGCTTCGAGGC 1540

QY 142 LysLys 143  
Db 1541 AGAAG 1546

RESULT 32  
US-09-285-310-1  
; Sequence 1, Application US/09285310  
; Patent No. 6262026  
; GENERAL INFORMATION:  
; APPLICANT: Heartlein, Michael W.  
; APPLICANT: Lemontt, Jeffrey F.  
; APPLICANT: Concino, Michael F.  
; TITLE OF INVENTION: CHIMERIC PROTEINS FOR USE  
; TITLE OF INVENTION: IN TRANSPORT OF A SELECTED SUBSTANCE INTO CELLS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; FILING DATE: US/09/285,310  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/037,188  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fraser, Ph.D., J.D., Janis K.  
; REGISTRATION NUMBER: 34,819  
; REFERENCE/DOCKET NUMBER: 07236/009002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154







QY 20 ProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyr 39  
Db 632 CCTAAAGGTTTAACTGAAGTACCGTTATGAATCGATGAATCGGAGCTAATAGCT 691  
QY 40 SerTyrLysTyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGly 59  
Db 692 TCGTTTGTCTTATACATCAT-----LysAspAsnSer 75  
QY 60 SerGlyTyrGluAlaPheThrLysLysGluGluTyr-----LysAspAsnSer 75  
Db 710 CAGGATATGATTTCTTATGCGAGTAAAGTTTGGTCATGCTGATTTACTACT 769  
QY 76 ArgPheAlaIleGlyThrTyrSerLeuAsnAspAlaLeuThrLeuArgAla----- 93  
Db 770 TCAGTAACAATGGGCCATCTTCCGCATCAACGAATATGTTAGCTTTATGATTACTG 829  
QY 94 GlyLeuAlaTyrAspLys-----AlaAlaSerLys 103  
Db 830 GGGGCGCTCATGGAAGGTTAAGCATCTGTATTGATGAATCAATCAGTCAAGAG 889  
QY 104 ThrHisLeuSerAlaSerIleProAspThrAspArgMetTyrTyrSerIleGlyAlaThr 123  
Db 890 AGCTCAATGGCA-----TAGGGCAGGGGTGCA 919  
QY 124 TyrLysPheThrProAsnLeuSerValAspValGlyPheAlaHis 138  
Db 920 TTCAACCCACTTCCAAATTTTCTCATGACGCTTATATGAATATAC 964  
RESULT 38  
US-08-628-434-1  
; Sequence 1, Application US/08628434  
; Patent No. 6265367  
; GENERAL INFORMATION:  
; APPLICANT: Sparling, P. Frederick  
; APPLICANT: Reucher, Margaret  
; TITLE OF INVENTION: Isolated FrpB Nucleic Acid Molecule and Vaccine  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Imclone Systems Incorporated  
; STREET: 180 Varick Street  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10014  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/628,434  
; FILING DATE: 05-APR-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/418,964  
; FILING DATE: 07-APR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gallagher, Thomas C.  
; REGISTRATION NUMBER: 37,066  
; REFERENCE/DOCKET NUMBER: SPA-3-P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 645-1405  
; TELEFAX: (212) 645-2054  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2600 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 318..2456

US-08-628-434-1  
Alignment Scores:  
Pred. No.: 3,73 Length: 2600  
Score: 77.50 Matches: 41  
Percent Similarity: 33.83% Conservative: 27  
Best Local Similarity: 20.40% Mismatches: 76  
Query Match: 8.10% Indels: 57  
DB: 4 Gaps: 10  
US-09-995-493-52 (1-179) x US-08-628-434-1 (1-2600)  
QY 7 GlyProTyrIleGlyLysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSer 26  
Db 1647 GCACAAACCGTTTCAAGCAGCAGCTCAACCCGAGTTTCGGCGTGATTGG----- 1697  
QY 27 GlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLysTyr----- 43  
Db 1698 -----CAGCGCGCGCAACACTGGAGCTTCAGCGCGAGCCACAACTACCGCGCGCG 1748  
QY 43 ----- 43  
Db 1749 AGCCCGCGCTGTATGACGCTCTGCAAAACCCAGCGGCGCATCATCTCGATTGGC 1808  
QY 44 -----ThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGly 59  
Db 1809 GACGCGCAGAAAGCGCGCGCGCATACCGAAATCGGCTTCAACTACACGACGCG 1868  
QY 60 SerGlyTyrGluAlaPheThrLysLysGluGluTyr-----LysAspAsnSerArgPheAla 78  
Db 1869 ACG-----TTTGGCGCAACACGCGCAGCTACTTCCGCGCAGACCATCAAGACGCG 1916  
QY 79 IleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAsp 98  
Db 1917 CTTGCCAATCCCAAAACCGCCGACCTCCGTCGCGCGCGGAA-----GCCGTCAAC 1970  
QY 99 LysAlaAlaSerLysThrHisLeuSerAlaSerIleProAspThrAspArgMetTyrTrp 118  
Db 1971 GCCGCGTACATCAAAACCCAC-----GGTTAC 1997  
QY 119 SerIleGlyAlaThrTyrLysPheThrProAsnLeuSerValAspValGlyPheAlaHis 138  
Db 1998 GAATTGGCGCGCTCTCTACCGC---ACCGCGCGCGCTGACCGCCAAAGTCGGCGTAAGCCAC 2054  
QY 139 LeuArgGlyLysLysLysHisPheValGluThrGlnAsnIleLysGlyLeuLeuVal 158  
Db 2055 -----AGCAACCGCGCTTTACGATACGACACAAAGACAG---CTGTTGACGCGG 2102  
QY 159 GluAlaAspTyrThrLysAlaThrAlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArg 178  
Db 2103 AACCCCTGAATTTGGCGCACAA---GTGCGCGCGCACTTGGACGCGCTCCCTTCCTACCCG 2159  
QY 179 Phe 179  
Db 2160 TTC 2162  
RESULT 39  
US-09-277-565-24/c  
; Sequence 24, Application US/09277565  
; Patent No. 6207384  
; GENERAL INFORMATION:  
; APPLICANT: Mekalanos, John J.  
; APPLICANT: Akerley, Brian J.  
; APPLICANT: Rubin, Eric J.  
; APPLICANT: Camilli, Andrew  
; TITLE OF INVENTION: SYSTEMATIC IDENTIFICATION OF ESSENTIAL  
; FILE OF INVENTION: GENES BY IN VITRO TRANSPOSON MUTAGENESIS  
; FILE REFERENCE: 00742/052002  
; CURRENT APPLICATION NUMBER: US/09/277,565  
; CURRENT FILING DATE: 1999-03-26  
; EARLIER APPLICATION NUMBER: 60/079,770  
; EARLIER FILING DATE: 1998-03-27  
; NUMBER OF SEQ ID NOS: 29

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 852
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
US-09-277-565-24

Alignment Scores:
Pred. No.: 0.885 Length: 852
Score: 77.00 Matches: 30
Percent Similarity: 45.54% Conservative: 16
Best Local Similarity: 29.70% Mismatches: 47
Query Match: 8.05% Indels: 8
DB: 4 Gaps: 2

US-09-995-493-52 (1-179) x US-09-277-565-24 (1-852)
QY 73 AspAsnSerArgPheAlaIleGlyThrTyrSerLeuAsnAspAlaLeuThrLeuArg 92
Db 494 GATTCATAAAAATCAAGTGGCGTCTTCTCGCGATATTTTCTAATTTATCTTCGCT 435
QY 93 AlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIleProAsp 112
Db 434 GCTTTACTTTTCGAGGTAAAGCCCTCAAAACAAAATCTATCAGAACGCAATCCCGAT 375
QY 113 ThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSerVal 132
Db 374 GCACAAAGACGGGTAAATPGCCGCAAGCTCCTGGCAAGGACACACTCGAATGCCAGCT 315
QY 133 AspValGlyPheAlaHisLeuArgGlyLysLysLys-----HisPheVal 147
Db 314 TCA-----CGCATTCGGCGTACTAAATGAACACCGGTCAGTAAATGGCGTC 264
QY 148 GluThrGlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThrLysAlaThr 167
Db 263 CCCGATCAGAAATCAAGGCAATATTACTCCCTCGTGTAGCTTTTCCACCAAAATATGC 204
QY 168 Ala 168
Db 203 GCT 201

RESULT 40
US-09-087-277-1
; Sequence 1, Application US/09087277B
; Patent No. 6169226
; GENERAL INFORMATION:
; APPLICANT: EK, BO
; APPLICANT: KHOSNOODI, Jamshid
; APPLICANT: LARSSON, Clas-Tomas
; APPLICANT: LARSSON, Hakan
; APPLICANT: RASK, Lars
; TITLE OF INVENTION: STARCH BRANCHING ENZYME II OF POTATO
; FILE REFERENCE: 003300-486
; CURRENT APPLICATION NUMBER: US/09/087,277B
; CURRENT FILING DATE: 1998-05-29
; EARLIER APPLICATION NUMBER: PCT/SE96/01558
; EARLIER FILING DATE: 1996-11-28
; EARLIER APPLICATION NUMBER: SE 9504272-7
; EARLIER FILING DATE: 1995-11-29
; EARLIER APPLICATION NUMBER: SE 9601506-0
; EARLIER FILING DATE: 1996-04-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3074
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: beII gene
; OTHER INFORMATION: (branching enzyme II) from Solanum tuberosum
; OTHER INFORMATION: (potato)
; FEATURE:
; NAME/KEY: CDS

; LOCATION: (189)..(2825)
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (189)..(332)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (333)..(2825)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (92)..(2156)
; OTHER INFORMATION: Nucleotides 92, 285, 1406, 1430, 1897 and 2156 are
; OTHER INFORMATION: n wherein n = A, C, G or T.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (285)..(287)
; OTHER INFORMATION: Amino acid -16 is Xaa wherein Xaa = Ile, Leu, Val
; OTHER INFORMATION: or Phe.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1404)..(1406)
; OTHER INFORMATION: Amino acid 358 is Xaa wherein Xaa = Leu or Phe.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1428)..(1430)
; OTHER INFORMATION: Amino acid 366 is Xaa wherein Xaa = Thr.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1896)..(1898)
; OTHER INFORMATION: Amino acid 522 is Xaa wherein Xaa = Tyr, Ser, Cys
; OTHER INFORMATION: or Phe.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2154)..(2156)
; OTHER INFORMATION: Amino acid 608 is Xaa wherein Xaa = Pro.
US-09-087-277-1

Alignment Scores:
Pred. No.: 7.4 Length: 3074
Score: 76.00 Matches: 16
Percent Similarity: 56.52% Conservative: 10
Best Local Similarity: 34.78% Mismatches: 18
Query Match: 7.94% Indels: 2
DB: 4 Gaps: 1

US-09-995-493-52 (1-179) x US-09-087-277-1 (1-3074)
QY 38 HistTyrSerTyrLysTyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGln 57
Db 765 CACCTTGATTACAGGTATTTCACAGTACAGAAACTGAGGAGGAGCAATTGACAAGTATGAG 824
QY 58 AspGlySerGlyTyrGluAlaPheThrLysLysGluLysLysAspAsnSerArgPhe 77
Db 825 -----GGTGGTTTGGAAAGCTTTTCTCGTGTATGAAAAAATGGGTTTCACTCGTAGT 878
QY 78 AlaIleGlyThrTyr 83
Db 879 GCTACAGGTATCACTTAC 896

Search completed: May 19, 2003, 19:58:38
Job time : 65 secs
```



GenCore version 5.1.4\_p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 19, 2003, 19:53:07 ; Search time 99 Seconds  
(without alignments)  
2332.312 Million cell updates/sec

Title: US-09-995-493-52

Perfect score: 957

Sequence: 1 OHNGVLGPGYIGKSLTLKP.....ADYTTKATANLYGLNLYRF 179

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 810007 seqs, 644969091 residues

Total number of hits satisfying chosen parameters: 1620014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh  
-MODEL=frame+p2n.model -runat\_12052003\_091311\_23115/app.query.fasta\_1.327  
-O=/cgn2\_1/USPTO\_spool/US0995493/runat\_12052003\_091311\_23115/app.query.fasta\_1.327  
-DB=PublishedApplications\_NA -GFM=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US0995493 @CGN\_1.1.96 -runat\_12052003\_091311\_23115  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MWAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PublishedApplications\_NA:

```
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	549	57.4	1598	10	US-09-452-599-26
2	133.5	13.9	1371	10	US-09-997-664-103
3	133.5	13.9	1562	10	US-09-997-664-91
4	87	9.1	23451	9	US-10-114-170-173

5	83.5	8.7	784	10	US-09-765-272-151
6	83	8.7	9057	9	US-10-114-170-194
7	81.5	8.5	1359	10	US-09-815-242-4086
8	80.5	8.4	4605	9	US-10-080-505-10
9	80.5	8.4	4822	9	US-10-080-505-14
10	80	8.4	547	9	US-09-854-133-169
11	80	8.4	547	10	US-09-738-973-169
12	80	8.4	2097	9	US-09-826-025-25
13	80	8.4	2347	10	US-09-880-107-3040
14	79.5	8.3	1992	9	US-10-098-841-310
15	78	8.2	493	10	US-09-867-550-289
16	78	8.2	1204	10	US-09-881-752A-187
17	77.5	8.1	554	9	US-10-066-543-2097
18	77	8.0	852	10	US-09-815-242-7171
19	76.5	8.0	9047	10	US-09-070-927A-12
20	76	7.9	2160	9	US-09-938-842A-628
21	76	7.9	3074	9	US-10-254-534-1
22	74	7.7	849	12	US-10-062-994-5
23	74	7.7	849	12	US-10-062-994-5
24	73.5	7.7	5253	9	US-10-080-943-3
25	73.5	7.7	5253	10	US-09-880-107-1655
26	73.5	7.7	7304	9	US-10-114-170-174
27	73	7.6	1022	10	US-09-850-351A-33
28	73	7.6	15261	9	US-09-764-891-10186
29	73	7.6	124884	9	US-09-913-514-1
30	73	7.6	125157	9	US-09-913-514-2
31	72.5	7.6	1254	10	US-09-925-297-58
32	72.5	7.6	1380	10	US-09-969-347-209
33	72	7.5	1239	10	US-09-974-300-4357
34	72	7.5	4305	9	US-10-080-505-8
35	71.5	7.5	807	9	US-09-938-842A-212
36	71	7.4	636	10	US-09-974-300-1094
37	71	7.4	1548	9	US-10-114-170-86
38	70.5	7.4	1365	10	US-09-822-849A-513
39	70.5	7.4	1464	9	US-09-738-626-216
40	70.5	7.4	1585	10	US-09-765-205-11
41	70.5	7.4	1631	10	US-09-864-864-301
42	70.5	7.4	1638	9	US-09-905-291A-260
43	70.5	7.4	1638	9	US-09-902-853-260
44	70.5	7.4	1638	9	US-09-907-824-260
45	70.5	7.4	1638	9	US-09-907-841-260

#### ALIGNMENTS

RESULT 1  
US-09-452-599-26  
; Sequence 26, Application US/09452599  
; Patent No. US20020055101A1  
; GENERAL INFORMATION:  
; APPLICANT: Bergeron, Michel G.  
; APPLICANT: Ouellette, Marc  
; APPLICANT: Roy, Paul H.  
; TITLE OF INVENTION: Specific and Universal Probes and Amplification Primers  
; TITLE OF INVENTION: To Rapidly Detect and Identify Common Bacterial  
; TITLE OF INVENTION: Pathogens and Antibiotic Resistance Genes from Clinical  
; TITLE OF INVENTION: Specimens for Routine Diagnosis in Micro  
; FILE REFERENCE: 12287.31  
; CURRENT APPLICATION NUMBER: US/09/452,599  
; CURRENT FILING DATE: 1999-12-01  
; PRIOR APPLICATION NUMBER: 08/526,840  
; PRIOR FILING DATE: 1995-09-11  
; PRIOR APPLICATION NUMBER: 08/304,732  
; PRIOR FILING DATE: 1994-09-12  
; NUMBER OF SEQ ID NOS: 177  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 26  
; LENGTH: 1598  
; TYPE: DNA  
; ORGANISM: Haemophilus influenzae  
US-09-452-599-26  
Alignment Scores:

Pred. No.: 8,54e-65 Length: 1598  
Score: 549.00 Matches: 105  
Percent Similarity: 76.74% Conservative: 27  
Best Local Similarity: 61.05% Mismatches: 34  
Query Match: 57.37% Indels: 6  
DB: 10 Gaps: 2

US-09-995-493-52 (1-179) x US-09-452-599-26 (1-1598)

Qy 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu 31  
Db 939 AAAGTAAATTACCTTTACATGCGGAGATACCTGAGAACTTCTGCTTTCCATCAATTA 998  
Qy 32 ThrAspGlnTrpAlaLeuHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51  
Db 999 ACTGACAACTTCAGCGGATATAGTTATATAATATACCATGGAGCTCGTTTAAACAAA 1058  
Qy 52 LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyr 71  
Db 1059 TTACATCCAGCTTCGAAGATGGTAAA-----AAAGCTTTTGTATAAAGAAATTAACAATAC 1112  
Qy 72 LysAspAsnSerArgPheAlaLeuGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeu 91  
Db 1113 AGTAATACCTCTCGTTGCTATAGGGCAAGTTATATCTTTATGAAAATTTGACCTTA 1172  
Qy 92 ArgAlaGlyLeuAlaTyrAspLysAlaLeuSerLysThrHisLeuSerAlaSerIlePro 111  
Db 1173 CGTGGGGTATGCTTACGATCAAGCGGCATCTCGTCATCACCGTAGTGTGCAATTTCCA 1232  
Qy 112 AspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131  
Db 1233 GATACCCATCGCACTTGGTATAGTTAGTGGCAACCTTATAAATTCAGCGCAATTTATCT 1292  
Qy 132 ValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsn 151  
Db 1293 GTTGATCTTGGCTATGCTTACTTAAAGGCAAAAGAGTTTACITTAAGAGAGTAAARACA 1352  
Qy 152 IleLysGly-----LeuLeuLeuValGluAlaAspTyrThrThrLysAlaThr 167  
Db 1353 ATAGGTGACAAAGCTACATTCGACATTCGAATACAACTGCAAAATATATCTTCTCAAGCACAC 1412  
Qy 168 AlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
Db 1413 GCAAACTTTTACGGTTGCAATTTAAATATATAGTTTC 1448

## RESULT 2

US-09-997-664-103  
; Sequence 103, Application US/09997664  
; Patent No. US20020151003A1  
; GENERAL INFORMATION:  
; APPLICANT: Ben-Bassat, Arle  
; APPLICANT: Cattermole, Monica  
; APPLICANT: Gatenby, Anthony A.  
; APPLICANT: Gibson, Katherine J.  
; APPLICANT: Ramos-Gonzalez, Isabel  
; APPLICANT: Ramos, Juan  
; APPLICANT: Sarlasani, Sima  
; TITLE OF INVENTION: Method for the Production of p-Hydroxybenzoate in Species of  
; TITLE OF INVENTION: Pseudomonas and Agrobacterium  
; FILE REFERENCE: BC1018 US CIP  
; CURRENT APPLICATION NUMBER: US/09/997,664  
; CURRENT FILING DATE: 2001-11-28  
; PRIOR APPLICATION NUMBER: 09/585,174  
; PRIOR FILING DATE: 2000-06-01  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 103  
; LENGTH: 1371  
; TYPE: DNA  
; ORGANISM: Pseudomonas mendocina KR-1

US-09-997-664-103

## Alignment Scores:

Alignment Scores:  
Pred. No.: 3,45e-08 Length: 1562  
Score: 133.50 Matches: 40  
Percent Similarity: 44.00% Conservative: 26  
Best Local Similarity: 26.67% Mismatches: 75

Pred. No.: 2,85e-08 Length: 1371  
Score: 133.50 Matches: 40  
Percent Similarity: 44.00% Conservative: 26  
Best Local Similarity: 26.67% Mismatches: 75  
Query Match: 13.95% Indels: 9  
DB: 10 Gaps: 4

US-09-995-493-52 (1-179) x US-09-997-664-103 (1-1371)

Qy 13 GlySerLeuThrLeuLys-----LeuProAlaTyrTrpGluLeuSerGlyPheHis 29  
Db 871 GGGGATATCCGCGTAAAGACTTCGAGATGCCGCCAGTCTGAGCTTGGCTTTCGCTCAT 930  
Qy 30 GlnLeuThrAspGlnTrpAlaLeuHisTyrSerTyrLysTyrThrGluTrpSerArg--- 48  
Db 931 CAATTCAACGAGCGCTTGGCTGCTGCTGATGCTCAAGCGTCTACTGGAGCGATGTC 990  
Qy 49 PheLysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLys 68  
Db 991 ATGGAAGACATCATGTGTGGATTTCAAAATCGCAGTCAGGTGGATTTGATATCGAATTACCA 1050  
Qy 69 GluGluTyrLysAspAsnSerArgPheAlaLeuGlyThrThrTyrSerLeuAsnAspAla 88  
Db 1051 CACAACATATCAGGATATACGGTGGCTTCCATCGCCACCGCTTACAGAGTTATGACAAG 1110  
Qy 89 LeuThrLeuArgAlaGlyLeuAlaTyrAspLysAlaLeuSerLysThrHisLeuSerAla 108  
Db 1111 CTAACCTCTCGTGTGATATAGCTATGCGCAACAGCGCTGGACACAGCTGATATTG 1170  
Qy 109 Ser---IleProAspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThr 127  
Db 1171 CCAGTAAATTCAGCTTATTTGAAGAAACAGCTTTCTCTCGTAGCGGATTTATAGTTTGTAT 1230  
Qy 128 Pro-----AsnLeuSerValAspValGlyPheAlaHisLeuArgGlyLysLys 143  
Db 1231 AAAAATCAAACTCAATTTGGCGATTTCTTTTGGCCTAAAGAGAGCTTGAACACACCA 1290  
Qy 144 LysHisPheValGluThrGlnAsnIleLys 153  
Db 1291 TCATACCTAAGCGCGCACCGAAGCTTGAAG 1320

## RESULT 3

US-09-997-664-91  
; Sequence 91, Application US/09997664  
; Patent No. US20020151003A1  
; GENERAL INFORMATION:  
; APPLICANT: Ben-Bassat, Arle  
; APPLICANT: Cattermole, Monica  
; APPLICANT: Gatenby, Anthony A.  
; APPLICANT: Gibson, Katherine J.  
; APPLICANT: Ramos-Gonzalez, Isabel  
; APPLICANT: Ramos, Juan  
; APPLICANT: Sarlasani, Sima  
; TITLE OF INVENTION: Method for the Production of p-Hydroxybenzoate in Species of  
; TITLE OF INVENTION: Pseudomonas and Agrobacterium  
; FILE REFERENCE: BC1018 US CIP  
; CURRENT APPLICATION NUMBER: US/09/997,664  
; CURRENT FILING DATE: 2001-11-28  
; PRIOR APPLICATION NUMBER: 09/585,174  
; PRIOR FILING DATE: 2000-06-01  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 91  
; LENGTH: 1562  
; TYPE: DNA  
; ORGANISM: Pseudomonas mendocina KR-1

US-09-997-664-91

Alignment Scores:  
Pred. No.: 3,45e-08 Length: 1562  
Score: 133.50 Matches: 40  
Percent Similarity: 44.00% Conservative: 26  
Best Local Similarity: 26.67% Mismatches: 75

Query Match: 13.95% Indels: 9  
DB: 10 Gaps: 4  
US-09-995-493-52 (1-179) x US-09-997-664-91 (1-1562)  
QY 13 GlySerLeuThrLeuLys-----LeuProAlaTyrTrpGluLeuSerGlyPheHis 29  
DB 1062 GCGGATATCCGGCTAAAGACTTCGAGATCCGCCGCGAGCTGACGTTCCGCTTTGCTCAT 1121  
QY 30 GlnLeuThrAspGlnTrpAlaLeuHisTyrSerTyrLysTyrThrGluTrpSerArg--- 48  
DB 1122 CATTCAACGAGCGTTGGCTGGTCTGCTGATGTCACGGCTCTACTGGAGCGATGTC 1181  
QY 49 PheLysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLys 68  
DB 1182 ATGGAAGACATCAGTCGATTTCAAATCCGACTCAGGTGGGATTGCATATCGAATACCA 1241  
QY 69 GluGluTyrLysAspAsnSerArgPheAlaIleGlyThrTyrSerLeuAsnAspAla 88  
DB 1242 CACAATATCAGGATATTACGGTGGCTCCATCGGCCGCTTACAGAGTTAATGACAAG 1301  
QY 89 LeuThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAla 108  
DB 1302 CTAACCTCTCGTCTGATATGCTATGCGCACAGCGCTGCAGTAGGCTGATATTCG 1361  
QY 109 Ser---IleProAspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThr 127  
DB 1362 CCAGTAATCCAGCTTATTGGAAGAACACAGCTTCTCTCGGTAGCAGCTATAGTTTGTAT 1421  
QY 128 Pro-----AsnLeuSerValAspValGlyPheAlaHisLeuArgGlyLysLys 143  
DB 1422 AAAAAATCAAACTCAATTGGCGATTCTTTGGCTAAAGAGAGAGCTTGAACACACCA 1481  
QY 144 LysHisPheValGluThrGlnAsnIleLys 153  
DB 1482 TCATACCTAAGCGCGCACCGNAACGTTGAAG 1511

RESULT 4

US-10-114-170-173/c  
; Sequence 173, Application US/10114170  
; Publication No. US20030023075A1  
; GENERAL INFORMATION:  
; APPLICANT: Blattner, Frederick R.  
; Perna, Nicole T.  
; Plunkett, Guy  
; Welch, Rod  
; TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53701-2113  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 8.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/114,170  
; FILING DATE: 01-Apr-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/453,702  
; FILING DATE: 03-DEC-1999  
; APPLICATION NUMBER: 60/110,955  
; FILING DATE: 04-DEC-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J.  
; REGISTRATION NUMBER: 27386

REFERENCE/DOCKET NUMBER: 960296.95017  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 251-5000  
TELEFAX: (608) 251-9166  
INFORMATION FOR SEQ ID NO: 173:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23451  
TYPE: nucleic acid  
TOPOLOGY: linear  
STRAINEDNESS: double  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 173:  
US-10-114-170-173  
Alignment Scores:  
Pred. No.: 3.95 Length: 23451  
Score: 87.00 Matches: 49  
Percent Similarity: 36.67% Conservative: 28  
Best Local Similarity: 23.33% Mismatches: 63  
Query Match: 9.09% Indels: 70  
DB: 9 Gaps: 12  
US-09-995-493-52 (1-179) x US-10-114-170-173 (1-23451)  
QY 9 TyrIleGlyLysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPhe 28  
DB 6332 TATGTTATCAGTGGTTCA-----AGACTGGCTGATAAT 6300  
QY 29 HisGlnLeuThrAspGlnTrpAla----- 36  
DB 6299 CATACCTATCGGCTGCTTATGCTCAGAGTAAGTACAGACTTTAAGAAATATAAGGA 6240  
QY 37 IleHisTyrSerTyrLysTyrThrGluTrp-----SerArgPhe 49  
DB 6239 GTGAATCTCCAATATCGATAT---GAGTGGGATTCTCCGCTCAGTCTCGTGGGAGTTT 6183  
QY 50 LysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAla-----PheThr 66  
DB 6182 TCGTATATGAAGTGTATTGGCGGACAGTCATCGGATGAAGTCAGCATTTTACAGA 6123  
QY 67 LysLysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThrTyrThrTyrSerLeuAsn 86  
DB 6122 CATCAGGCTGATATTAATATTAATTCAC---TTTCTCGCAGGCCAGCTTATCGTCTGAAT 6066  
QY 87 AspAlaLeuThrLeu-----ArgAlaGlyLeuAlaTyrAspLysAla----- 100  
DB 6065 GACTATATCAGCTTTTATGCTTTGCTGGGATATCTCATACTAAACGAGGAGATTAT 6006  
QY 101 -----AlaSerLysThrHisLeuSerAlaSerIle--- 110  
DB 6005 GACTGGAGAAACAGCTGCGGTGCTGATGAGTCTGATGGATATTAAGTGTCTGTATCA 5946  
QY 111 ---ProAspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsn 129  
DB 5945 AAAAAATCCACTGAT---TTTGCTTATGACGCTGGTGTATTATTATTCGCTGGGCAAT 5889  
QY 130 LeuSerValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThr 149  
DB 5888 ATCTCTGTAATGTTGGGTAT-----GAGGGAACATAA----- 5856  
QY 150 GlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsn 169  
DB 5855 -----GCAGATATATATGAAAAACATTCGGTCAAT 5826  
QY 170 LeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
DB 5825 GGATTACTGTTGGAGTTGGTTATCGTTTT 5796  
RESULT 5  
US-09-765-272-151  
; Sequence 151, Application US/09765272  
; Patent No. US20020061545A1  
; GENERAL INFORMATION:

```

; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272.
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 151:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 784 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 151:
US-09-765-272-151

Alignment Scores:
Pred. No.: 0.0819 Length: 784
Score: 83.50 Matches: 42
Percent Similarity: 39.53% Conservatives: 26
Best Local Similarity: 24.42% Mismatches: 77
Query Match: 8.73% Indels: 28
DB: 10 Gaps: 7

US-09-995-493-52 (1-179) x US-09-765-272-151 (1-784)
Qy 8 ProTyrIleGlyLysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGly 27
Db 23 CCGTTTCATCTAAGGTAGTACAACTGGTCCCTACCCAAATGGTATCAGATTAGAAGGT 82
Qy 28 PheHisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLys-----TyrThr 44
Db 83 TTTCCAAAG-----TCAGAGTGG-----TACTACTTCGATAAAATGGAGTCTACAA 130
Qy 45 GluTrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAla 64
Db 131 CAGTTTGTGGTTGGAAACATTAGAGATTAAACTAAAGACAGTGTGGGAAGAAAGTAC 190
Qy 65 PheThrLysLysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThrTyrSer 84
Db 191 GGGGAAAACGTGAAGATTTCAGAGATTAAGAGAGAGCGCTTATTATACCACTATTAC 250
Qy 85 LeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAla----- 100
Db 251 TTTAATCAAAATCATCTCTTAGACACAGGTGGCTTTATGATCAGTCTAAGTGGTATTAT 310
Qy 101 AlaSerLysThrHisLeuSer-----AlaSerIle 110
Db 311 CTAGCTAAGACGGAATTAATTAATGGAGAAACTACCTTGGTGTGAAGACGTCGGCGGTGG 370
Qy 111 ProAspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeu 130

```

```

Db 371 ATAAACGATGATTCGACTTGTGTACTACTAGATCCAAACAACGATGATTT----- 418
Qy 131 SerValAspValGlyPheAlaHisLeuArgGlyLysLysHisPheValGluThrGln 150
Db 419 ---ATGCAACAGAGTTGGCAATATCTA---GGTAATAGTGGTACTACCTCCGTTCTCTCA 472
Qy 151 AsnIleLysGlyLeuLeuValGluAlaAspTyr 162
Db 473 GGAGC-AATGCCACTGGCTGATCAGGAAGGTAC 507

RESULT 6
US-10-114-170-194/c
; Sequence 194, Application US/10114170
; Publication No. US20030023075A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. US20030023075A1e1 Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114,170
; FILING DATE: 01-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/453,702
; FILING DATE: 03-DEC-1999
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 194:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9057
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 194:
US-10-114-170-194

Alignment Scores:
Pred. No.: 3.45 Length: 9057
Score: 83.00 Matches: 42
Percent Similarity: 37.02% Conservatives: 25
Best Local Similarity: 23.20% Mismatches: 58
Query Match: 8.67% Indels: 56
DB: 9 Gaps: 10

US-09-995-493-52 (1-179) x US-10-114-170-194 (1-9057)
Qy 33 AspGlnTrpAlaIleHis-----TyrSerTyrLysTyrThrGluTrpSerArgPheLys 50

```

Db 1897 GACAAATGGTCATCTCGTCGGGGGATGACTATCAATCCGACTAATCGGCTGATG----- 1844  
QY 51 GluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGlu 70  
Db 1843 ---TTATTTGGCTCATAT-----GCCAGGCATTCGGCGCCCGACGATGGCGAAATG 1793  
QY 71 TyrLysAspAsnSerArgPheAlaIleGlyThrTyrSer----- 84  
Db 1792 TATAACGATTCTAAGCACTTCTCGATTGGTCGCTTCTATACCAACTATTGGGTGCCAAAC 1733  
QY 84 ----- 84  
Db 1732 CCGAACTTACGTCGCGAAACTACGAACTCAGGAGTACGGTTTGGGCTGCGCTTTTGAT 1673  
QY 85 -----LeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAla 100  
Db 1672 GACCTGATGTTGTCCAATGATGCTCTGGAATTTAAAGCCAGCTACTTTGATACCAAGCG 1613  
QY 101 -----AlaSerLysThrHisLeuSerAlaSerIle 110  
Db 1612 AAGGATTACATCTCCACGACCGTCGATTTTCGGCGCGCGACGACTATGTCGTATAACGTC 1553  
QY 111 ProAspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeu 130  
Db 1552 CCGAACGCC---AAATCTGG---GCGTGGGATGTGATGACGAAATATACCACTGATCTG 1499  
QY 131 ---SerValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThr 149  
Db 1498 TTTAGCTTGATGTCGCTATACCTACCGTACCGCGCGGCAAA-----GACACC 1454  
QY 150 GlnAsnIleGlyLysLeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsn 169  
Db 1453 GATACCGCGGAATACATCTCCAGCATTAACCCGAT---ACTGTACCACTCTGAAT 1397  
QY 170 Leu 170  
Db 1396 ATT 1394

RESULT 7

US-09-815-242-4086  
; Sequence 4086, Application US/09815242  
; Patent No. US20030061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; FILE REFERENCE: ELITRA 011A  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows version 4.0  
; SEQ ID NO 4086

; LENGTH: 1359  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-4086  
Alignment Scores:  
Pred. No.: 0.343 Length: 1359  
Score: 81.50 Matches: 35  
Percent Similarity: 40.67% Conservative: 26  
Best Local Similarity: 23.33% Mismatches: 56  
Query Match: 8.52% Indels: 33  
DB: 10 Gaps: 7  
US-09-995-493-52 (1-179) x US-09-815-242-4086 (1-1359)  
QY 4 GlyValLeuGlyProTyrIleGlyLysGlySerLeuThrLeuLysLeuProAlaTyr--- 22  
Db 703 GGCTAGTCGGCGCTACACCTTCGCCCGACGACCTGCGCAACGCTCGCCGCTACGAA 762  
QY 23 -----TrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyrSer 40  
Db 763 AGCCACTTCAAGCAGACGCGC---AAGCAGCTGATACCGCTGCTCGCCGCC 819  
QY 41 TyrLysTyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGlySer 60  
Db 820 ATCGGCTTACCAGGAA-----TCGCTGTGGCAGCCGCGGCC 855  
QY 61 GlyTyrGluAlaPheThrLysLysGluGluTyrLys-----AspAsn 74  
Db 856 -----ACCTCCAAGACCGCGCTGCGCGCTGATGCTGACCAACCGG 900  
QY 75 SerArgPheAlaIleGlyThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGly 94  
Db 901 ACCGCCAGCGCGATGGCGGTCTCCACCGCTCGACCCGAGCAGACGATCCAGGCGGC 960  
QY 95 LeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIleProAspThrAsp 114  
Db 961 AGCAAGTAT---TTCGTGCAGATCCCGCAGCAACTGCGCCGAGAGCATCAAGAACCGGAC 1017  
QY 115 ArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSerValAspVal 134  
Db 1018 CGCAGCTGGTTCGCCCTCGCGCC-----TACACATC 1050  
QY 135 GlyPheAlaHisLeuArgGlyLysLysLys 144  
Db 1051 GCGCGCGCGCACCTGGGAAGACGCGCGCAAG 1080

RESULT 8

US-10-080-505-10  
; Sequence 10, Application US/10080505  
; Publication No. US20030073166A1  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme, Joseph W.  
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS  
; FILE REFERENCE: A-59941-1/FT/DCF/DHR  
; CURRENT APPLICATION NUMBER: US/10/080,505  
; CURRENT FILING DATE: 2002-02-22  
; PRIOR APPLICATION NUMBER: US 08/296,791  
; PRIOR FILING DATE: 1994-10-25  
; PRIOR APPLICATION NUMBER: US 09/839,996  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 4605  
; TYPE: DNA  
; ORGANISM: Haemophilus influenzae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (422)..(4597)  
; OTHER INFORMATION:  
US-10-080-505-10

Alignment Scores:  
 Pred. No.: 2.8 Length: 4605  
 Score: 80.50 Matches: 34  
 Percent Similarity: 38.9% Conservativeness: 28  
 Best Local Similarity: 21.38% Mismatches: 51  
 Query Match: 8.41% Indels: 46  
 DB: Gaps: 7

US-09-995-493-52 (1-179) x US-10-080-505-10 (1-4605)

QY 22 TyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyr 41  
 DB 994 CATGTGAGACTTACCTATGCGAGTTTCATGGTTAAATAGTGG-----AAATACCTTTGA 1047  
 QY 42 LysTyrThrGluTrpSerArgPhe-LysGluLeuArgGlyLysTyrGlnAspGlySerG1 61  
 DB 1048 AGATGACCAGCTGGTAAACGGTACATTAGAAATTAATGGCGAGTACAAATCCTAATAA 1107  
 QY 61 yTyrGluAlaPheThrLysLysGluGluTyrLysAspAsnSerArgPheAlaIleGlyTh 81  
 DB 1108 ATATGCTCCACTACCTACGCGAGTTTCATTCGGGGATAGT-----GGTTC 1152  
 QY 81 rThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAla 101  
 DB 1153 TCCAATGTTTATTT-----TATGATAAGGAAGT 1179  
 QY 101 aserLysThrHisLeuSerAlaSerIleProAspThrAspArgMetTrpTyrSerIleG1 121  
 DB 1180 TAAGAAATGGTTTAAATGCGCGTGTACGTGAA---GGAAATCCTTATGCTGCAGTAGG 1236  
 QY 121 yAlaThrTyrLysPheThrProAsnLeuSerValaspValGlyPheAlaHisLeuArgG1 141  
 DB 1237 AACAGCTATCAAAATTACA-----CAAGGTATTCTT----- 1255  
 QY 141 yLysLysLysHisPheValGluThrGlnAsnIleLysGlyLeuLeuValGluAlaAs 161  
 DB 1256 -CGAAAGATTATTT-----CAAGGTATTCTT----- 1282  
 QY 161 pTyrThrThrLysAlaThrAlaAsnLeuTyrGlyLeuAsnTyrArgPhe 179  
 DB 1283 ----AATCAAGACATTACAGCTAAATTTTGGGATACTAATGCTGAATATAGATT 1333

RESULT 9  
 US-10-080-505-14  
 ; Sequence 14, Application US/10080505  
 ; Publication No. US20030073166A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: St. Geme, Joseph W.  
 ; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS  
 ; FILE REFERENCE: A-59941-1/RFT/DCE/DHR  
 ; CURRENT APPLICATION NUMBER: US/10/080,505  
 ; CURRENT FILING DATE: 2002-02-22  
 ; PRIOR APPLICATION NUMBER: US 08/296,791  
 ; PRIOR FILING DATE: 1994-10-25  
 ; PRIOR APPLICATION NUMBER: US 09/839,996  
 ; PRIOR FILING DATE: 2001-04-20  
 ; NUMBER OF SEQ ID NOS: 58  
 ; SOFTWARE: Patent version 3.1  
 ; SEQ ID NO 14  
 ; LENGTH: 4822  
 ; TYPE: DNA  
 ; ORGANISM: Haemophilus influenzae  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (388)..(4563)  
 ; OTHER INFORMATION:  
 ; US-10-080-505-14

Alignment Scores:  
 Pred. No.: 3 Length: 4822  
 Score: 80.50 Matches: 34  
 Percent Similarity: 38.9% Conservativeness: 28  
 Best Local Similarity: 21.38% Mismatches: 51

Query Match: 8.41% Indels: 46  
 DB: Gaps: 7  
 US-09-995-493-52 (1-179) x US-10-080-505-14 (1-4822)  
 QY 22 TyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyr 41  
 DB 960 CATGTGAGACTTACCTATGCGAGTTTCATGGTTAAATAGTGG-----AAATACCTTTGA 1013  
 QY 42 LysTyrThrGluTrpSerArgPhe-LysGluLeuArgGlyLysTyrGlnAspGlySerG1 61  
 DB 1014 AGATGACCAGCTGGTAAACGGTACATTAGAAATTAATGGCGAGTACAAATCCTAATAA 1073  
 QY 61 yTyrGluAlaPheThrLysLysGluGluTyrLysAspAsnSerArgPheAlaIleGlyTh 81  
 DB 1074 ATATGCTCCACTACCTACGCGAGTTTCATTCGGGGATAGT-----GGTTC 1118  
 QY 81 rThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAla 101  
 DB 1119 TCCAATGTTTATTT-----TATGATAAGGAAGT 1145  
 QY 101 aserLysThrHisLeuSerAlaSerIleProAspThrAspArgMetTrpTyrSerIleG1 121  
 DB 1146 TAAGAAATGGTTTAAATGCGCGTGTACGTGAA---GGAAATCCTTATGCTGCAGTAGG 1202  
 QY 121 yAlaThrTyrLysPheThrProAsnLeuSerValaspValGlyPheAlaHisLeuArgG1 141  
 DB 1203 AACAGCTATCAAAATTACA-----CAAGGTATTCTT----- 1221  
 QY 141 yLysLysLysHisPheValGluThrGlnAsnIleLysGlyLeuLeuValGluAlaAs 161  
 DB 1222 -CGAAAGATTATTT-----CAAGGTATTCTT----- 1248  
 QY 161 pTyrThrThrLysAlaThrAlaAsnLeuTyrGlyLeuAsnTyrArgPhe 179  
 DB 1249 ----AATCAAGACATTACAGCTAAATTTTGGGATACTAATGCTGAATATAGATT 1299

RESULT 10  
 US-09-854-133-169  
 ; Sequence 169, Application US/09854133  
 ; Publication No. US20020183499A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Mohamath, Raodoh  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Secrist, Heather  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
 ; THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
 ; FILE REFERENCE: 210121.475C10  
 ; CURRENT APPLICATION NUMBER: US/09/854,133  
 ; CURRENT FILING DATE: 2001-05-11  
 ; NUMBER OF SEQ ID NOS: 735  
 ; SOFTWARE: FastSeq for Windows version 3.0  
 ; SEQ ID NO 169  
 ; LENGTH: 547  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien  
 ; US-09-854-133-169

Alignment Scores:  
 Pred. No.: 0.145 Length: 547  
 Score: 80.00 Matches: 34  
 Percent Similarity: 40.16% Conservativeness: 15  
 Best Local Similarity: 27.87% Mismatches: 53  
 Query Match: 8.36% Indels: 20  
 DB: Gaps: 6

US-09-995-493-52 (1-179) x US-09-854-133-169 (1-547)  
 QY 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51  
 DB 124 ACTGTGATGGTGTGTCAGTGTGCGAGCATGAGGCCACCTAAGTCCCGGAC 183







```
Query Match: 8.31% Indels: 31
DB: 9 Gaps: 6
US-09-995-493-52 (1-179) x US-10-098-841-310 (1-1992)
QY 30 GluLeuThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPhe 49
Db 965 AAAATCGAGACTATCTTTGTCACATTCAGTAACCGAGAAGATGCAGTTGAGGCTATG 1024
QY 50 LysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGlu 69
Db 1025 AAAGCTTTAAATGGCAAGGTGCTGGATGGTTCCCGCATTCACCTACCCCTAGCAAAACCA 1084
QY 70 GluTyrLysAspAsn---SerArgPheAlaIleGlyThr----- 81
Db 1085 GTGGACAAGGACAGATTATGTAGTATACCCGAGGCACAGGTGGAAGGGCCACCATCTG 1144
QY 82 -----ThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyr 97
Db 1145 CAAGGAGAGTATACCTACTCTTTG-----GGCCAAGTTTAT 1180
QY 98 AspLysAlaAlaSerLysThrHisLeuSerAlaSerIleProAspThrAspArgMetTrp 117
Db 1181 GAT-----CCCAACCAACCTACCTTGGAGCTCCTGTC----- 1213
QY 118 TyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSerValAspValGlyPheAla 137
Db 1214 TTCTATGCCCCCAGACCTATGAGCAATTCGCCAGTCTTCATTTCCAGCCACCAAGGA 1273
QY 138 HisLeuArgGlyLysLysHisPheValGluThrGlnAsnIleLysGly 154
Db 1274 CATCTC-----AGCAACAGAGCCATTATCCGAGCCCTTCTCTGTAGAGG 1318
RESULT 15
US-09-867-550-289
; Sequence 289, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehrahan, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: Nos. US20020082206A1 Polynucleotides from Atherogenic Cells and
; FILE REFERENCE: Thereby
; FILE REFERENCE: 21402-013 (Curra-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 289
; LENGTH: 493
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)
; OTHER INFORMATION: wherein any n is one of a or t or g or c
US-09-867-550-289
Alignment Scores:
Pred. No.: 0.233 Length: 493
Score: 78.00 Matches: 37
Percent Similarity: 36.55% Conservative: 16
Best Local Similarity: 25.52% Mismatches: 56
Query Match: 8.15% Indels: 36
DB: 10 Gaps: 6
US-09-995-493-52 (1-179) x US-09-867-550-289 (1-493)
QY 17 LeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAla 36
```

```
Db 33 CTGAACACCTCCGCGCCTGCTCCAGGTGCAACACCTGCTCTTCGGCTCG--- 89
QY 37 IleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyr 56
Db 90 -----TGGTACGCGCTACACGGTGCATGAGGAACTGTCGGACTTCCGCGCGTCACC 140
QY 57 GlnAspGlySerGlyTyrGluAlaPheThrLysLysGlu----- 69
Db 141 TCGGACGGCAACGACTATCTCGGCTACAGCTCCGAGTGATCGCGTGGCGGCAAGATC 200
QY 70 -----GluTyrLysAspAsnSerArgPheAlaIleGlyThr 81
Db 201 CTCTCGGCTACGACCTCAACTTCATGCAGCAGGTCAACACGCTGTTTCATCGCGCACAC 260
QY 82 ThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAla 101
Db 261 CAGCGCGTCTCAACGACAGGCTCACCCCTCAGCGCGGCTTC-----CGCGCGGCC 311
QY 102 SerLysThrHisLeuSerAlaSer---IleProAspThrAspArgMetTrpTyrSerIle 120
Db 312 ATGCTCTCGCGCCGAGCCACCACTCATCCCGCGCGCGCATCCC-----TACAAGATG 365
QY 121 -----GlyAlaThrTyrLysPheThrProAsn 129
Db 366 CAGGCGCAACTACTCGAGCGCTCGCGCAGTTCTCGCGCCAGTACCAAGATCACCCCGCAC 425
QY 130 LeuSerValAspVal 134
Db 426 GATCAGGTGTTCGTC 440
RESULT 16
US-09-881-752A-187
; Sequence 187, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleenathous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Comen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in
; FILE REFERENCE: Genome
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 187
; LENGTH: 1204
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (51)...(1151)
US-09-881-752A-187
Alignment Scores:
Pred. No.: 0.862 Length: 1204
Score: 78.00 Matches: 35
Percent Similarity: 42.42% Conservative: 21
Best Local Similarity: 26.52% Mismatches: 46
Query Match: 8.15% Indels: 30
DB: 10 Gaps: 7
US-09-995-493-52 (1-179) x US-09-881-752A-187 (1-1204)
QY 63 GluAlaPheThrLysLysGluTyrLysAspAsnSerArg---PheAlaIleGlyThr 81
Db 456 GAACCCATTACTAACCCCTTTAGAAATTTGAAGAAACTTAAAGAAATTTAGAAAGTCAATTT 515
```



Db 374 GCACAAAGAGCGGTAATTCGCCACAAAGCTCTGCAAAAGGCACAACTCGAATGCCAGCT 315  
Qy 133 AspValGlyPheAlaHisLeuArgGlyLysLysLys-----HisPheVal 147  
Db 314 TCA-----CGCATGGCGTACTAATAAGAAACACAGGCTCACTGATTATGCCGTC 264  
Qy 148 GluThrGlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThrThrLysAlaThr 167  
Db 263 CCCGCATCAGAAATCAAGGCAATATTACTCCCTGCTTGGAGCTTTTCCACCAAAATATGC 204  
Qy 168 Ala 168  
Db 203 GCT 201

RESULT 19  
US-09-070-927A-12  
; Sequence 12, Application US/09070927A  
; Patent No. US20020120116A1  
; GENERAL INFORMATION:  
; APPLICANT: Charles A. Kunsch  
; Patrick J. Dillon  
; Steven Barash  
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 982  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/070,927A  
; FILING DATE: 04-May-2000  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/046,655  
; FILING DATE: 1997-05-16  
; APPLICATION NUMBER: 60/044,031  
; FILING DATE: 1997-05-06  
; APPLICATION NUMBER: 60/066,009  
; FILING DATE: 1997-11-14

ATTORNEY/AGENT INFORMATION:  
; NAME: Kenley K. Hoover  
; REGISTRATION NUMBER: 40,302  
; REFERENCE/DOCKET NUMBER: PB369  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9047 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear

US-09-070-927A-12  
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
Alignment Scores:  
Pred. No.: 26.4 Length: 9047  
Score: 76.50 Matches: 41  
Percent Similarity: 37.09% Conservative: 15  
Best Local Similarity: 27.15% Mismatches: 45  
Query Match: 7.99% Indels: 50  
DB: 10 Gaps: 6

US-09-995-493-52 (1-179) x US-09-070-927A-12 (1-9047)

Qy 18 LysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAlaIle 37  
Db 3630 AAAATTCAGCC-----AAGTTAAGTTTCTATTATTAGCGATTACAGAC----- 3671  
Qy 38 HisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGln 57  
Db 3672 -----AGTCGTTTGCCTGTTATGATTCT 3695  
Qy 58 AspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyrLysAspAsnSerArgPhe 77  
Db 3696 GATAATGCTGGGTTTGATGCTTTTCAAAAGAGCGGCACAAAGGCCCTATCAAGACTTT 3755  
Qy 78 AlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGly-----Leu 95  
Db 3756 TTAACAGGC-----AAATTGACCTTCCCAACTGCTCAACAGCTA 3794  
Qy 96 AlaTyrAspLysAlaAlaSerLysThrHisLeuSer----- 107  
Db 3795 GCAGAGGATAGTTAGCTCTCTAAATTTGTTTATGCAATGTTTTCGAAGTCAAAACATGAG 3854  
Qy 108 -----AlaSerIleProAspThrAspArgMet 116  
Db 3855 TCGATTGAGGTCAACAGACAGCGGTCAGGCCCTTAACTACACAGAAAAACAATATT 3914  
Qy 117 TrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSerValValGly-Ph 136  
Db 3915 TTTTATTCGGATGGTTCACACGAGAGTGAAGTCAATGCTCA-----GAAGTT 3965  
Qy 136 eAlaHisLeuArgGlyLysLysHisPhe 146  
Db 3966 CCAGCATCTTCTTATCAAAAAGAGGAATAT 3996

RESULT 20  
US-09-938-842A-628  
; Sequence 628, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; FILE REFERENCE: SCRIPI300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-08-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 628  
; LENGTH: 2160  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-628

Alignment Scores:  
Pred. No.: 3.8 Length: 2160  
Score: 76.00 Matches: 52  
Percent Similarity: 28.43% Conservative: 31  
Best Local Similarity: 24.07% Mismatches: 73  
Query Match: 7.94% Indels: 60  
DB: 9 Gaps: 14

US-09-995-493-52 (1-179) x US-09-938-842A-628 (1-2160)  
Qy 3 AsnGly-----ValLeuGlyProTyrIleGlyLysGlySerLeuThrLeuLysLeuPro 20  
Db 151 AATGGAGCCACCGTCTTCTTCGCTATGTTGTCGATCAGCG-----AAAAAGCTTCCT 204



```

; APPLICANT: Whitmire, William M.
; APPLICANT: Allemen, Arthur R.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and
; FILE REFERENCE: UF-167XC3
; CURRENT APPLICATION NUMBER: US/10/062,994
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/533,662
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 09/337,827
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 08/953,326
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 08/733,230
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 849
; TYPE: DNA
; ORGANISM: Anaplasma marginale
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(846)
; US-10-062-994-5

Alignment Scores:
Pred. No.: 1.81 Length: 849
Score: 74.00 Matches: 49
Percent Similarity: 38.00% Conservative: 27
Best Local Similarity: 24.50% Mismatches: 86
Query Match: 7.73% Indels: 38
DB: 12 Gaps: 11

US-09-995-493-52 (1-179) x US-10-062-994-5 (1-849)

QY 14 SerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPhe----- 28
Db 250 ACCGATTGATGTAGTGTGCAGCAAACTTTCCAAATCTGGCTACACTTTTGCTTCTCT 309
QY 29 ---HisGlnLeuThrAsp---GlnTrpAlaIleHisTyrSerTyr----- 41
Db 310 AAAAAGCTTAATCAGCTTTTCGACGGCGCTGTGGGATATCTCTGGGAGGAGCCAGAGTG 369
QY 42 ---LysTyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGlySer 60
Db 370 GAATTGGAAGCGAGCTACAGAGTTTGTCTACTTTGGCGGCGGAGTACGCAAAAGT 429
QY 61 GlyTyrGlu-----AlaPheThrLysLysGluGluTyrLysAspAsnSerArgPhe 77
Db 430 GGTGCGGAATCTCTGGCAGCTATTACCGCGAGCTAACATTACTGAGACCAATTACTTC 489
QY 78 AlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyr 97
Db 490 GTAGTCAAAATGATGAATCAACAACCTCAGTCAGTATGTTAAATGGC-----TGCAT 543
QY 98 AspLysAlaAlaSerLysThrHisLeuSer-----AlaSer. 109
Db 544 GAGCTGCTGCACACAGATTTTACCTGTGTCCTCCCGTATGTATGTCGGGATAGCGCAAGC 603
QY 110 IleProAsp-----ThrAspArgMetTrpTyr-----SerIleGlyAla 122
Db 604 TTTGTTGACATCTCTAAGCAAGTAACACAAAGCTGCGCTACAGGGGCAAGGTTGGGATT 663
QY 123 ThrTyrLysPheThrProAsnLeuSer---ValAspValGlyPheAlaHis-----Leu 139
Db 664 AGCTACAGTTTACTCCGGAATATCTCTGGTGGCAGGTTGGTTCTACCAAGCGGCTATTT 723
QY 140 ArgGlyLysLysLysHisPheValGluThrGlnAsnIleLysGlyLeuLeuValGlu 159
Db 724 GATGAGCTTTACAGGACATCTCCCGCACACACAGCTGAAG-----TTCTCTGGAGAA 777
QY 160 AlaAspTyrThrThrLysAlaThrAlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179
```

```

Db 778 GCAAAGCCTCAGTCAAGCGCATATTTGCTGACTAGGCTTTAACTTTGGAGCAAGATTC 837
RESULT 23
US-10-062-994-5
; Sequence 5, Application US/10062994
; Patent No. US20020132789A1
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Bowle, Michael V.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Mahan, Suman M.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Moreland, Annie L.
; APPLICANT: Simbi, Bigboy H.
; APPLICANT: Whitmire, William M.
; APPLICANT: Allemen, Arthur R.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and
; FILE REFERENCE: UF-167XC3
; CURRENT APPLICATION NUMBER: US/10/062,994
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/533,662
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 09/337,827
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 08/953,326
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 08/733,230
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 849
; TYPE: DNA
; ORGANISM: Anaplasma marginale
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(846)
; US-10-062-994-5

Alignment Scores:
Pred. No.: 1.81 Length: 849
Score: 74.00 Matches: 49
Percent Similarity: 38.00% Conservative: 27
Best Local Similarity: 24.50% Mismatches: 86
Query Match: 7.73% Indels: 38
DB: 12 Gaps: 11

US-09-995-493-52 (1-179) x US-10-062-994-5 (1-849)

QY 14 SerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPhe----- 28
Db 250 ACCGATTGATGTAGTGTGCAGCAAACTTTCCAAATCTGGCTACACTTTTGCTTCTCT 309
QY 29 ---HisGlnLeuThrAsp---GlnTrpAlaIleHisTyrSerTyr----- 41
Db 310 AAAAAGCTTAATCAGCTTTTCGACGGCGCTGTGGGATATCTCTGGGAGGAGCCAGAGTG 369
QY 42 ---LysTyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGlySer 60
Db 370 GAATTGGAAGCGAGCTACAGAGTTTGTCTACTTTGGCGGCGGAGTACGCAAAAGT 429
QY 61 GlyTyrGlu-----AlaPheThrLysLysGluGluTyrLysAspAsnSerArgPhe 77
Db 430 GGTGCGGAATCTCTGGCAGCTATTACCGCGAGCTAACATTACTGAGACCAATTACTTC 489
QY 78 AlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyr 97
Db 490 GTAGTCAAAATGATGAATCAACAACCTCAGTCAGTATGTTAAATGGC-----TGCAT 543
QY 98 AspLysAlaAlaSerLysThrHisLeuSer-----AlaSer 109
```

```
Db 544 GACGTGTCACACAGATTTACCTGTCGCCGCTATGATGTCGCCGGGATAGCGCAAGC 603
Qy 110 IleProasp-----ThrAspArgMetTrpTyr-----SerIleGlyAla 122
Db 604 TTGTGTCACATCTCTAGCAAGTAACCAACAACTGCGCTACAGGGCGCAAGTTGGGATT 663
Qy 123 ThrTyrLysPheThrProAsnLeuSer---ValAspValGlyPheAlaHis-----Leu 139
Db 664 AGCTACCAGTTTACTCGGAAATATCTTGTGTCAGGTGGTCTTCTACCACGGCTATTT 723
Qy 140 ArgGlyLysLysHisPheValGluThrGlnAsnIleLysGlyLeuLeuValGlu 159
Db 724 GATGAGCTTACAGGACATTCGCGCACACACAGTGTAAAG-----TTCTCTGGAGAA 777
Qy 160 AlaAspTyrThrThrLysAlaThrAlaAsnLeuTyrGlyLeuAsnLeuAsnTyrAtqPhe 179
Db 778 GCANAACCTCAGTCAAGCGCATATTGCTGACTACGGCTTTAACTTTGGAGCAGATTC 837
RESULT 24
US-10-080-943-3
; Sequence 3, Application US/10080943
; Patent No. US20030073236A1
; GENERAL INFORMATION:
; APPLICANT: Field, Loren J.
; APPLICANT: Tsai, Shih-Chong
; TITLE OF INVENTION: P193 PROTEINS AND NUCLEIC ACIDS, AND USES THEREOF
; FILE REFERENCE: IU99-PC7
; CURRENT APPLICATION NUMBER: US/10/080,943
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: 60/150,266
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 5253
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (87)..(5183)
US-10-080-943-3
Alignment Scores:
Pred. No.: 30.6 Length: 5253
Score: 73.50 Matches: 28
Percent Similarity: 38.46% Conservative: 17
Best Local Similarity: 23.93% Mismatches: 47
Query Match: 7.68% Indels: 25
DB: 9 Gaps: 3
US-09-995-493-52 (1-179) x US-10-080-943-3 (1-5253)
Qy 64 AlaPheThrLysLysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThrThrTyr 83
Db 1134 TCCTTCAGGAGGTCAAGACGCTTTTCGCCCTCGCTTCAGTTCGCAAGTGCAATACCTAT 1193
Qy 84 SerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLys 103
Db 1194 GCITTTGATGTCGGGGACACACTGCACCGGGGATGCGAGTGGGATGATTAT 1253
Qy 104 ThrHisLeuSerAla-----SerIleProasp 112
Db 1254 GAGGAGATCAGTCGCCGGGATGAGGGCGAGTTTCGGCAGACAAACACGGTGTGCTCCT 1313
Qy 113 ThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSerVal 132
Db 1314 GTGCAGGTATTTGGGAGTCAACAGCCGCCACCTATTGGGTGGCAGCTGGCGGAG 1373
Qy 133 AspValGlyPheAlaHisLeuArgGlyLysLysHisPheValGluThrGlnAsnIle 152
Db 1374 ATCTTGGGCTTT-----GAGGAAGACATT 1397
US-10-114-170-174
; Sequence 174, Application US/10114170
; Publication No. US20030023075A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
```

```
Qy 153 LysGlyLeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsn 169
Db 1398 GAGGAC-----ATGGTTGAGGCTGATGAGTACCAAGGGCGAGTGGCCAGT 1442
RESULT 25
US-09-880-107-1655
; Sequence 1655, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1655
; LENGTH: 5253
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D38548
US-09-880-107-1655
Alignment Scores:
Pred. No.: 30.6 Length: 5253
Score: 73.50 Matches: 28
Percent Similarity: 38.46% Conservative: 17
Best Local Similarity: 23.93% Mismatches: 47
Query Match: 7.68% Indels: 25
DB: 10 Gaps: 3
US-09-995-493-52 (1-179) x US-09-880-107-1655 (1-5253)
Qy 64 AlaPheThrLysLysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThrThrTyr 83
Db 1134 TCCTTCAGGAGGTCAAGACGCTTTTCGCCCTCGCTTCAGTTCGCAAGTGCAATACCTAT 1193
Qy 84 SerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLys 103
Db 1194 GCITTTGATGTCGGGGACACACTGCACCGGGGATGCGAGTGGGATGATTAT 1253
Qy 104 ThrHisLeuSerAla-----SerIleProasp 112
Db 1254 GAGGAGATCAGTCGCCGGGATGAGGGCGAGTTTCGGCAGACAAACACGGTGTGCTCCT 1313
Qy 113 ThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSerVal 132
Db 1314 GTGCAGGTATTTGGGAGTCAACAGCCGCCACCTATTGGGTGGCAGCTGGCGGAG 1373
Qy 133 AspValGlyPheAlaHisLeuArgGlyLysLysHisPheValGluThrGlnAsnIle 152
Db 1374 ATCTTGGGCTTT-----GAGGAAGACATT 1397
Qy 153 LysGlyLeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsn 169
Db 1398 GAGGAC-----ATGGTTGAGGCTGATGAGTACCAAGGGCGAGTGGCCAGT 1442
RESULT 26
US-10-114-170-174
; Sequence 174, Application US/10114170
; Publication No. US20030023075A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
```

```

;
; Plunkett, Guy
; weich, Rod
;
; TITLE OF INVENTION: No. US20030023075A1e1 Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Plinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114,170
; FILING DATE: 01-Apr-2002
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/453,702
; FILING DATE: 03-DEC-1999
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
;
; INFORMATION FOR SEQ ID NO: 174:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7304
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 174:
;
; US-10-114-170-174
;
; Alignment Scores:
; Pred. No.: 49.6 Length: 7304
; Score: 73.50 Matches: 40
; Percent Similarity: 38.71% Conservative: 32
; Best Local Similarity: 21.51% Mismatches: 65
; Query Match: 7.68% Indels: 49
; DB: 9 Gaps: 12
;
; US-09-995-493-52 (1-179) x US-10-114-170-174 (1-7304)
;
; QY 16 ThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGln-----Leu 31
; |||:||||| :||| :||| :|||
; Db 6358 ACCATCAACAGGAGAGCTAC-----TCAGCTATGACAGGATCGGCTCTATCTG 6408
;
; QY 32 ThrAspGlnTrp-----AlaIleHisTyrSerTyrLysTyrThrGluTrpSer 47
; ||||| ||| ||| ||| |||||:|||||
; Db 6409 ACCGATAACTGGATTGCGGTCGCGGATCCGCTATCAGTAT--TACAGCGCAGTATGCG 6465
;
; QY 48 ArgPheLysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLys 67
; ||||| ||||| ||| :|||
; Db 6466 -----GGCTAA-----GGCGGTCCTTTTATGTCAATACTGAC 6498
;
; QY 68 LysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAsp 87
; :||| :||| :||| :|||
; Db 6499 AGCCGGATGAACATGAGCGCCCAA-----CTGGGTATGCTACAAACTGACGCCA 6552
;
; QY 88 AlaLeuThrLeuArgAlaGlyLeuAla-----TyrAspLysAlaAlaSerLysThrHis 105
; :|||:||||| ||| :|||
; Db 6553 TCGGTATCCTTATTGCTAATTATTCGCAACATTTATGCGCAATCGTCAATTCGCAGC 6612
;
; QY 106 LeuSerAlaSerIleProAspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLys 125

```

```

;
; Db 6613 TACATCGGAGATCTTCACCGGAATCATCTAATGCTTACGAAGTCGGGCAAAATTCGAG 6672
; :|||:||||| :||| :|||:|||||
; QY 126 PheThrProAsnLeuSerValAspValGlyPheAlaHisLeuArgGlyLysLysHis 145
; :|||:||||| :||| :|||:|||||
; Db 6673 CTGTTGATGATATAACCGCAGATATTGCGCTGTTTGTATATCCAT-----AAACGTTAAC 6726
; :|||:||||| :||| :|||:|||||
; QY 146 PheValGluThrGlnAsnIle----- 152
; :|||:||||| :||| :|||:|||||
; Db 6727 GTGTTGTATACCGAAAGTATTGGTGATGAACCATCGCCAAACGGCAGCGCGTTTCGT 6786
; :|||:||||| :||| :|||:|||||
; QY 153 ---LysGlyLeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeu--- 170
; :|||:||||| :||| :|||:|||||
; Db 6787 TCAAGAGGG-----GTAGAAGTCGACCTTTCGGGGAGCATTAACGTGAACATTAAAT 6837
; :|||:||||| :||| :|||:|||||
; QY 171 -----TyrGly 172
; :|||:||||| :||| :|||:|||||
; Db 6838 ATCATTTGCCAGTACGGC 6855
; :|||:||||| :||| :|||:|||||
;
; RESULT 27
; US-09-850-351A-33/C
; Sequence 33, Application US/09850351A
; Patent No. US20020100080A1
; GENERAL INFORMATION:
; APPLICANT: Feitelson, Jerald S.
; NAME: Schnepf, H. Ernest
; Narva, Kenneth E.
; Stockhoff, Brian A.
; Schmeits, James
; Loewer, David
; Dullum, Charles Joseph
; Muller-Cohn, Judy
; Stamp, Lisa
; Morfill, George
;
; TITLE OF INVENTION: No. US20020100080A1e1 Pesticidal Toxins and Nucleotide
; Sequences Which Encode These Toxins
;
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/850,351A
; FILING DATE: 07-May-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/073,898
; FILING DATE: 06-MAY-1998
; APPLICATION NUMBER: US 08/960,780
; FILING DATE: 30-OCT-1997
; APPLICATION NUMBER: US 60/029,848
; FILING DATE: 30-OCT-1996
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanders, Jay M.
; REGISTRATION NUMBER: 39,355
; REFERENCE/DOCKET NUMBER: MA-708CD1
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
;
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1022 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: 17718  
SEQUENCE DESCRIPTION: SEQ ID NO: 33:  
US-09-850-351A-33

Alignment Scores:  
Pred. No.: 3, 26 Length: 1022  
Score: 73.00 Matches: 35  
Percent Similarity: 38.40% Conservative: 13  
Best Local Similarity: 28.00% Mismatches: 49  
Query Match: 7.63% Indels: 29  
DB: 10 Gaps: 7

US-09-995-493-52 (1-179) x US-09-850-351A-33 (1-1022)

```
Qy 29 HisGlnLeuThrAspGlnTrp---AlaIleHisTyrSerTyrLysTyrThrGluTrp--- 46
Db 878 CACGAGCTT-----CAATGGAGAGCTCTTCTGATTTCGTATAGACCAATTCGTGGATG 825
Qy 47 SerArgPheLysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThr 66
Db 824 AATGAGAC-TCTACACTATTGGATAAATTTTCATTGGT----- 787
Qy 67 LysLysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThrTyrSerLeuAsn 86
Db 786 -----GATAATATCACCTTTTCCATA---CTAACATTCACACTTGGGR 748
Qy 87 AspAlaLeuThr-----LeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThr 104
Db 747 AAGCAGCTACCAATGGGTGACGTTTCCCTTTCGATTTGATAAATCTAATCCCTTTGGCG 688
Qy 105 HisLeuSerAlaSerIleProAspThrAspArgMetTyrSerIleGlyAlaThrTyr 124
Db 687 GCCTTTTCATATCAGTAGGATGCGCAACTGCTGGCTGCTAATGGATTCGAACA 628
Qy 125 LysPheThrProAsnLeuSerValAspValGlyPheAlaHisLeuArgGlyLysLys 144
Db 627 AATTTTCTGA-----TATCCCTTACTTGTAGCGAATCATCC 592
Qy 145 HisPheValGluThr 149
Db 591 CATTTGACGCAACT 577
```

## RESULT 28

US-09-764-891-10186/c  
Sequence 10186, Application US/09764891  
Publication No. US20030077808A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC006  
CURRENT APPLICATION NUMBER: US/09/764,891  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 10231  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 10186  
LENGTH: 15261  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-891-10186

Alignment Scores:  
Pred. No.: 170 Length: 15261  
Score: 73.00 Matches: 49  
Percent Similarity: 33.98% Conservative: 21  
Best Local Similarity: 23.79% Mismatches: 81  
Query Match: 7.63% Indels: 55  
DB: 9 Gaps: 10

US-09-995-493-52 (1-179) x US-09-764-891-10186 (1-15261)

```
Qy 12 LysGlySerLeuThrLeuLysLeuProAlaTyrTrpGlu-LeuSerGlyPheHis----- 29
Db 10993 AAGCAATCCTCCTCCCTGGCCTCCCAAGTGTGGGAATCAGCGCTGAGCCACTGTGC 10934
Qy 30 -----GlnLeuThrAspGlnTrpAlaIleHisTyrSe 40
Db 10933 CCAGCCCAATGCTAATAACATGAAGAACCCCACTGAGTGACCATTCCTTAGACACTATCA 10874
Qy 40 rTyrLysTyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGlySe 60
Db 10873 TGTCAACTTTCAGACCTTGGGAGACACAGAAA-----AACAAATATAAGAT----- 10828
Qy 60 rGlyTyrGluAlaPheThrLysLysGluGluTyrLysAspAsnSerArgPheAlaIleG 80
Db 10827 -----AGAATTACAGGCAAGAAGACGTAATCCTGGGGCAAGGACCTTCCTCAA 10778
Qy 80 yThrThrTyrSerLeuAsn----- 86
Db 10777 CACTTCTTAACAAACCAAAATAGTGGCTTATTAGATGCATTTCTTGGCTGCAGAGGAAAAAT 10718
Qy 87 -----AspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAsp----- 98
Db 10717 GATAAGCAACTCACCTCCCATCTCAGATTGAGGGGTGTCTTGTGTAATGACAATTGTCT 10658
Qy 99 -----LysAlaAlaSerLysThrHisLeuSerAlaSerIlePro-----AspThrAs 114
Db 10657 TCCCAAGCTTGGATTCCCAAGTCCACCTTGTAGTCCCTGCGTAAAGCTCTCGCTTGA 10598
Qy 114 pArgMet---TrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSerValas 133
Db 10597 CAGCTGCATCGTGGTTTCTCGAAACATTAATGGGGAAGAAAGTCCCTTCAATA----- 10546
Qy 133 pValGlyPheAlaHisLeuArgGlyLysLysLysPheValGluThrGlnAsnIleLy 153
Db 10545 -TTATAGTTCTCTCAACAGAAAGGAAAAAATCATGAATCCAAATTCAGGCTCTTAC 10487
Qy 153 sGlyLeuLeuVal-----GluAlaAspTyrThrLysAlaThrLysAlaAsnLeuTy 171
Db 10486 TCATTTCAAATCATCTCCTGAAAGGACACACAGACACTTAAAGTAGCAGCT----- 10435
Qy 171 rGlyLeuAsnLeuAsn 176
Db 10434 -GGGAGAATCTCAAC 10420
```

## RESULT 29

US-09-913-514-1  
Sequence 1, Application US/09913514  
Publication No. US20030082210A1  
GENERAL INFORMATION:  
APPLICANT: GOMI, Yasuyuki  
TITLE OF INVENTION: Method for Quality Control of an Attenuated Varicella Live Vac  
FILE REFERENCE: 0216-0454P  
CURRENT APPLICATION NUMBER: US/09/913,514  
CURRENT FILING DATE: 2001-12-07  
PRIOR APPLICATION NUMBER: PCT/Jp01/00678  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: JP 2000-62734  
PRIOR FILING DATE: 2000-01-31  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 124884  
TYPE: DNA  
ORGANISM: Varicella virus  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)..(124884)  
OTHER INFORMATION: Dumas Strain  
US-09-913-514-1



## Alignment Scores:

Pred. No.: 3.7e+03 Length: 124884  
Score: 73.00 Matches: 41  
Percent Similarity: 38.51% Conservative: 16  
Best Local Similarity: 27.70% Mismatches: 42  
Query Match: 27.70% Indels: 50  
DB: 7.63% Gaps: 9

US-09-995-493-52 (1-179) x US-09-913-514-1 (1-124884)

QY 10 IleGlyLysGlySerLeu-----ThrLeuLysLeu-----ProAlaTyrTrp 23  
DB 39637 TTAGGGAACCCAGATCTTCTATAGAGCTTTGTTGGAAAAAATTAAACCTTTTACTTGG 39696  
QY 24 GluLeuSerGlyPheHis-----GlnLeu 31  
DB 39697 ATACTGATGGGTTCACCTTCGATATAGAGTGTGTATAAATCAGTTGGGGTTAAACTTG 39756  
QY 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGlu----- 45  
DB 39757 GAACCATGCTATGCAGTCAGGTGGATTATCTTTACGACCGGCACCTCCCGATGAAGGTA 39816  
QY 46 ---TrpSer-----ArgPheLysGluLeuArg 53  
DB 39817 TTGTGGAAGAACTACACTATCGGCATTACGGTGGCCCAATGAGGTC-AATGAGCTACGC 39875  
QY 54 GlyLysTyrGlnAspGly-----SerGlyTyrGluAlaPheThr----- 66  
DB 39876 ATTGAATACGAATCCGCTATAAATCCGGGTTTCTGCCTTTTCCACCTTTGTTAGGCAT 39935  
QY 67 ---LysLysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThrThrTyrSer--- 84  
DB 39936 CGCCACCGCGAATGGGTAAACCAACGACGACGAGCCATTGCAGAGATATACGCCGCG 39995  
QY 85 ---LeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLys-----Ala 100  
DB 39996 CTTATAACAAACACATTTGACAGCAATACGGGGTTTCATTGGGCAAGCTTATTATTCT 40055  
QY 101 AlaSerLysThrHisLeuSerAla 108  
DB 40056 TTTGAAAAACACCACCTTAATTCT 40079

## RESULT 30

US-09-913-514-2  
; Sequence 2, Application US/09913514  
; Publication No. US20030082210A1  
; GENERAL INFORMATION:  
; APPLICANT: GOMI, Yasuyuki  
; APPLICANT: SUNAMACHI, Hiroki  
; APPLICANT: TAKAHASHI, Michiaki  
; APPLICANT: YAMANISHI, Koichi  
; TITLE OF INVENTION: Method for Quality Control of an Attenuated Varicella Live Vaccine  
; FILE REFERENCE: 0216-0454P  
; CURRENT APPLICATION NUMBER: US/09/913,514  
; CURRENT FILING DATE: 2001-12-07  
; PRIOR APPLICATION NUMBER: PCT/JP01/00678  
; PRIOR FILING DATE: 2001-01-31  
; PRIOR APPLICATION NUMBER: JP 2000-62734  
; PRIOR FILING DATE: 2000-01-31  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 125157  
; TYPE: DNA  
; ORGANISM: Varicella virus  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(125157)  
; OTHER INFORMATION: Attenuated Oka strain  
US-09-913-514-2

## Alignment Scores:

Pred. No.: 3.71e+03 Length: 125157

Score: 73.00 Matches: 41  
Percent Similarity: 38.51% Conservative: 16  
Best Local Similarity: 27.70% Mismatches: 42  
Query Match: 27.70% Indels: 50  
DB: 7.63% Gaps: 9

US-09-995-493-52 (1-179) x US-09-913-514-2 (1-125157)

QY 10 IleGlyLysGlySerLeu-----ThrLeuLysLeu-----ProAlaTyrTrp 23  
DB 39588 TTAGGGAACCCAGATCTTCTATAGAGCTTTGTTGGAAAAAATTAAACCTTTTACTTGG 39647  
QY 24 GluLeuSerGlyPheHis-----GlnLeu 31  
DB 39648 ATACTGATGGGTTCACCTTCGATATAGAGTGTGTATAAATCAGTTGGGGTTAAACTTG 39707  
QY 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGlu----- 45  
DB 39708 GAACCATGCTATGCAGTCAGGTGGATTATCTTTACGACCGGCACCTCCCGATGAAGGTA 39767  
QY 46 ---TrpSer-----ArgPheLysGluLeuArg 53  
DB 39768 TTGTGGAAGAACTACACTATCGGCATTACGGTGGCCCAATGAGGTC-AATGAGCTACGC 39826  
QY 54 GlyLysTyrGlnAspGly-----SerGlyTyrGluAlaPheThr----- 66  
DB 39827 ATTGAATACGAATCCGCTATAAATCCGGGTTTCTGCCTTTTCCACCTTTGTTAGGCAT 39886  
QY 67 ---LysLysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThrThrTyrSer--- 84  
DB 39887 CGCCACCGCGAATGGGTAAACCAACGACGACGAGCCATTGCAGAGATATACGCCGCG 39946  
QY 85 ---LeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLys-----Ala 100  
DB 39947 CTTATAACAAACACATTTGACAGCAATACGGGGTTTCATTGGGCAAGCTTATTATTCT 40006  
QY 101 AlaSerLysThrHisLeuSerAla 108  
DB 40007 TTTGAAAAACACCACCTTAATTCT 40030

## RESULT 31

US-09-925-297-58  
; Sequence 58, Application US/09925297  
; Patent No. US20020081659A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA105  
; CURRENT APPLICATION NUMBER: US/09/925,297  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1993-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 58  
; LENGTH: 1254  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-297-58

Alignment Scores:  
Pred. No.: 5.14 Length: 1254  
Score: 72.50 Matches: 45  
Percent Similarity: 33.82% Conservative: 25  
Best Local Similarity: 21.74% Mismatches: 66  
Query Match: 21.74% Indels: 71  
DB: 7.58% Gaps: 10

US-09-995-493-52 (1-179) x US-09-925-297-58 (1-1254)

QY 18 LysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAlaIle 37

```
Db 517 AACGCTCAGCCATCTGGATCGACACGGCATCCAT-----TCCCGGGAGTGGGTCAACC 570
Qy 38 HistyrsertyrlystyrThrGluTrpSerArgPheHisGlnLeuThrAspGlnTrpAlaIle 37
Db 571 CAGGCCAGT-----GGGCTCTGG----- 588
Qy 58 AspGlySerGlyTyrGluAlaPheThrLysLys-----GluGluTyrLysAspAsnSer 75
Db 589 -----TTTGCAAGAAGATCACTCAAGACTACGGGCAGGATGCA 627
Qy 76 ArgPheAla-----IleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeuArgAla 93
Db 628 GCTTTTCCACGCCCATCTCGACACCTTGGACATCTTCTGGAGATCGTCACCAACCCCTGAT 687
Qy 94 GlyLeuAlaTyrAspLysAlaAla-----SerLysThrHisLeuSer 107
Db 688 GGCTTTCCACGCCCATCTCGACACCAAGATCGGATGGCGCAAGACTCGGTCCCACACAGCA 747
Qy 108 AlaSerIle-----ProAspThrAspArgMetTrpTyrSer----- 119
Db 748 GGCTCCCTCTGTATTGGCTGGACCCCAACAGGAACCTGGGACGCTGGCTTTGGGTTGTCC 807
Qy 120 -----IleGlyAlaThrTyr-----LysPheThrProAsnLeuSer 131
Db 808 GGAGCCAGCAGTAACCCCTGCTCGGAGACTTACCACGGCAAGTTTGCC---AATTCCGAA 864
Qy 132 ValAsp-----ValGlyPheAlaHisLeuArgGlyLysLysHisPheVal 147
Db 865 GTGGAGGTCAAGTCCATTGTAGACTTTGTGAGGACCATGGGAACATCAAGGCCTTCATC 924
Qy 148 GluThrGlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThrLys----- 165
Db 925 TCCATCCACAGCTACTCCAGCTCCTCATGATCCCTATGGCTACAAACAGAACAGCAGTC 984
Qy 166 -----AlaThrAlaAsnLeu 170
Db 985 CCGACACGAGTACGCTGGATGAGCTTTTCCAGGCTGCTGTGACAGCCCTGGCCTCTCTC 1044
Qy 171 TyrGlyLeuAsnLeuAsnTyr 177
Db 1045 TACGGGACCAAGTTCAACTAT 1065
```

## RESULT 32

```
US-09-969-347-209
; Sequence 209, Application US/09969347
; Patent No. US20020115085A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-69
; CURRENT APPLICATION NUMBER: US/09/969,347
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/60/237,598
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,604
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 318
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 209
; LENGTH: 1380
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-969-347-209
```

```
Alignment Scores:
Pred. No.: 5.92 Length: 1380
Score: 72.50 Matches: 45
Percent Similarity: 33.82% Conservative: 25
Best Local Similarity: 21.74% Mismatches: 66
Query Match: 7.58% Indels: 71
DB: 10 Gaps: 12
```

```
US-09-995-493-52 (1-179) x US-09-969-347-209 (1-1380)
Qy 18 LysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAlaIle 37
Db 509 AAGCGTCACGCCATCTGGATCGACACGGGCATCCAT-----TCCCGGGAGTGGGTCAACC 562
Qy 38 HistyrsertyrlystyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGln 57
Db 563 CAGGCCAGT-----GGGCTCTGG----- 580
Qy 58 AspGlySerGlyTyrGluAlaPheThrLysLys-----GluGluTyrLysAspAsnSer 75
Db 581 -----TTTGCAAGAAGATCACTCAAGACTATGGGCAGGATGCA 619
Qy 76 ArgPheAla-----IleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeuArgAla 93
Db 620 GCTTTTCCACGCCCATCTCGACACCTTGGACATCTTCTGGAGATCGTCACCAACCCCTGAT 679
Qy 94 GlyLeuAlaTyrAspLysAlaAla-----SerLysThrHisLeuSer 107
Db 680 GGCTTTCCCTTTCACGCACACAGCAATCGCATGGCGCAAGACTCGGTCCCACACAGCA 739
Qy 108 AlaSerIle-----ProAspThrAspArgMetTrpTyrSer----- 119
Db 740 GGCTCCCTCTGTATTGGCTGGACCCCAACAGGAACCTGGGACGCTGGCTTTGGGTTGTCC 799
Qy 120 -----IleGlyAlaThrTyr-----LysPheThrProAsnLeuSer 131
Db 800 GGAGCCAGCAGTAACCCCTGCTCGGAGACTTACCACGGCAAGTTTGCC---AATTCCGAA 856
Qy 132 ValAsp-----ValGlyPheAlaHisLeuArgGlyLysLysHisPheVal 147
Db 857 GTGGAGGTCAAGTCCATTGTAGACTTTGTGAGGACCATGGGAACATCAAGGCCTTCATC 916
Qy 148 GluThrGlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThrLys----- 165
Db 917 TCCATCCACAGCTACTCCAGCTCCTCATGATCCCTATGGCTACAAACAGAACAGCAGTC 976
Qy 166 -----AlaThrAlaAsnLeu 170
Db 977 CCTCACCAGGATGAGCTGGATGAGCTTTTCCAGGCTGCTGTGACAGCCCTGGCCTCTCTC 1036
Qy 171 TyrGlyLeuAsnLeuAsnTyr 177
Db 1037 TACGGGACCAAGTTCAACTAT 1057
```

## RESULT 33

```
US-09-974-300-4357
; Sequence 4357, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berk, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4357
; LENGTH: 1239
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-4357
```

```
Alignment Scores:
Pred. No.: 5.91 Length: 1239
```

Score: 72.00 Matches: 52  
Percent Similarity: 33.62% Conservative: 26  
Best Local Similarity: 22.41% Mismatches: 68  
Query Match: 7.52% Indels: 86  
DB: 10 Gaps: 13

US-09-995-493-52 (1-179) x US-09-974-300-4357 (1-1239)

QY 22 TyrTrpGluLeu-----SerGlyPheHis 29  
:|||||:|||||  
Db 478 TTCTGGAAGCTATTGACCGGGCACATTCGGAAGATCAGAAATCTCCACGCTTTCAC 537  
QY 30 GlnLeu----- 31  
:|||||  
Db 538 CAGCTTGGCATCGGAGTTCCTCAACTATTTCGGTCATCATCTCCGGCTGAAAGACGAA 597  
QY 32 -----ThrAspGlnTrpAlaHisTyrSerTyrLysTyrThrLutrpSerArg 48  
:|||||:|||||:|||||  
Db 598 ATCGAAGAAAAACCGAAAGCAGCTCCATTACTTGTCTGGAACGACGAGCAGCTGCAA 657  
QY 49 -----PheLysGluLeu----- 52  
:|||||  
Db 658 ATCTTCTGGCTACATCACTATATAGCTGATTGTCTCGCTCGCCAAAACGGGG 717  
QY 53 -----ArgGlyLys---TyrGlnAspGlySerGlyTyrGluAlaPheThrLysLys----- 68  
:|||||:|||||:|||||  
Db 718 CAAAAGGCGCGCTCAATGATTTAAAGCAATTTCCGCGCAATATACAAAAGCAACTT 777  
QY 69 ---GluGluTyrLys---AspAsnSerArgPheAlaHisGlyThrTyrSerLeuAsn 86  
:|||||:|||||:|||||  
Db 778 GAAGACGCGCTATAACTGAATACGTCAGCATCGCGGATCGCGGGAATGTACAGCTTG--- 834  
QY 87 AspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeu 106  
:|||||:|||||:|||||  
Db 835 -----ATCACCTT---GTACACCACTCTTATCAGGAAGCGCTCTCGCGCTGAAGGTC 885  
QY 107 SerAlaSerIleProAspThrArgMetTrpTyrSerIleGly----- 121  
:|||||:|||||  
Db 886 AAGGAGCGTTTCCCGCATGAGACAGGCAAGCCCTCGTCAGCTTTTCAGAGCTCGCACTAC 945  
QY 122 -----AlaThrTyrLysPheThrProAsn 129  
:|||||:|||||  
Db 946 CAGTATCTCGAGTCTCAGCGAAACGGAACATGCAAGCTAT-----CCAAAC 996  
QY 130 LeuSerValAspValGlyPheAlaHisLeuArgGlyLysLysHisPheValGluThr 149  
:|||||:|||||:|||||  
Db 997 TATCTTTAATCAGCTTCAAGCATATGATAAGAGCACCACCTCCCAATCTTGTGAGAGC 1056  
QY 150 GlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThrLysAlaThrAla--- 168  
:|||||:|||||  
Db 1057 CTTGAA-----CAATTTCATCGAATGCGACGACGACGTCACACCGCGCAAAA 1104  
QY 169 -----AsnLeuTyrClyLeuAsnLeuAsnTyrArg 178  
:|||||:|||||  
Db 1105 AAGCTGAACATTCACGCTCAACACGCTGAATTCACAG 1140

## RESULT 34

US-10-080-505-8  
; Sequence 8, Application US/10080505  
; Publication No. US20030073166A1  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme, Joseph W.  
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS  
; FILE REFERENCE: A-59941-1/RFT/DCP/DHR  
; CURRENT APPLICATION NUMBER: US/10/080,505  
; CURRENT FILING DATE: 2002-02-22  
; PRIOR APPLICATION NUMBER: US 08/296,791  
; PRIOR FILING DATE: 1994-10-25  
; PRIOR APPLICATION NUMBER: US 09/839,996  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO. 8

LENGTH: 4305  
TYPE: DNA  
ORGANISM: Haemophilus influenzae  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1702)..(1702)  
OTHER INFORMATION: "n" at position 1702 can be any base.  
NAME/KEY: CDS  
LOCATION: (1)..(4305)  
OTHER INFORMATION:  
US-10-080-505-8

## Alignment Scores:

Pred. No.: 36.6 Length: 4305  
Score: 72.00 Matches: 34  
Percent Similarity: 35.96% Conservative: 30  
Best Local Similarity: 19.10% Mismatches: 56  
Query Match: 7.52% Indels: 58  
DB: 9 Gaps: 9

US-09-995-493-52 (1-179) x US-10-080-505-8 (1-4305)

QY 46 TrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPhe 65  
:|||||:|||||:|||||  
Db 3565 TCGACAAATATATCGCACAGGATAAAGACGCTAT---GATTCTGATCGGTTCGCTGCTTAT 3621  
QY 66 ThrLysLysGluGluTyrLys-----AspAsnSerArgPhe 77  
:|||||:|||||  
Db 3622 CAGCAGAAACCAACTTGGCTCAATTTGGGTGCAAAAGCCCTTAGATTAATGACGAATT 3681  
QY 78 AlaIleGlyThrThrTyrSerLeuAsnAsp----- 87  
:|||||:|||||  
Db 3682 GGGCGGTTTTCTCGCATAGCCGTTTCAGATAATACCTTTGACGACAGGTTAAATAATCAC 3741  
QY 88 ---AlaLeuThrLeuArgAlaGlyLeuAlaTyr----- 97  
:|||||:|||||  
Db 3742 GCGACATTAAACGATGATGTCGGGTTTGGCCCAATATCAATGCGGCGATTACAAATTTGGT 3801  
QY 98 -----AspLysAlaAlaSerLysThr 104  
:|||||:|||||  
Db 3802 GTAAACGTCGGCGCGGAATTAGTCGAGTAAATGGCTGAAGAACAAAGCCGAAAAATT 3861  
QY 105 HisLeuSerAlaSerIleProAspThrArgMetTrpTyrSerIleGlyAlaThrTyr 124  
:|||||:|||||:|||||  
Db 3862 CATCGAAAAGCG-----ATAAATTATGCTGCTGAATGCAAGCTTAT 3900  
QY 125 LysPheThrProAsnLeuSerValAspValGlyPheAlaHisLeuArgGlyLysLys 144  
:|||||:|||||  
Db 3901 CAGTTCCCGTTTAGG-----CAATTGGGTATTTCAGCCTTATTGGGTGTTAATCGA 3951  
QY 145 HisPheValGluThrGlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThrThr 164  
:|||||:|||||:|||||  
Db 3952 TATTTTATGAAAGCGTGAATAATTATCAATCT-----GAAGAACTGAAGTGAACACA 4002  
QY 165 LysAlaThrAla---AsnLeuTyr-----GlyLeuAsnLeuAsnTyrArgPhe 179  
:|||||:|||||  
Db 4003 CCGAGCGCTTGCATTTAATCGCTATAATGCTGGCATTTCGAGTTGCATTATACATTT 4056

## RESULT 35

US-09-938-842A-212  
; Sequence 212, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; FILE REFERENCE: SAME, AND METHODS OF USE  
; TITLE OF INVENTION: SAME, AND METHODS OF USE  
; FILE REFERENCE: SCRIP1300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866

;; PRIOR FILING DATE: 2000-08-24  
;; PRIOR APPLICATION NUMBER: US 60/264,647  
;; PRIOR FILING DATE: 2001-01-16  
;; PRIOR APPLICATION NUMBER: US 60/300,111  
;; PRIOR FILING DATE: 2001-06-22  
;; NUMBER OF SEQ ID NOS: 5379  
;; SEQ ID NO 212  
;; LENGTH: 807  
;; TYPE: DNA  
;; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-212

Alignment Scores: Length: 807  
Pred. No.: 3.69  
Score: 71.50  
Percent Similarity: 39.13%  
Best Local Similarity: 22.98%  
Query Match: 7.47%  
DB: 9

US-09-995-493-52 (1-179) x US-09-938-842A-212 (1-807)

Qy 27 GlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrp 46  
Db 22 GGTTCCTTCAGTGTGTGAAGGCAAG---TCCAGCCATCCGATGGAGGACTATGTAGTG--- 75  
Qy 47 SerArgPheLysGluLeuArgGly----- 54  
Db 76 TCTGAATTCAGAAACCTTGAAGGTCATGAATTCGGTTTGTCTATCTTTGATGGTCAC 135  
Qy 55 -----LysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLys 68  
Db 136 TTGGGGCATGATGTGGCTAATAACTTCAGACTAATCTTTGCACACATCTTTAAAGAG 195  
Qy 69 GluGluTyrLysAspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAla 88  
Db 196 AAGATTTTTGGAGTACACAGAGATGCTATAGGAATCGCTACAGATCAACAGATGCC 255  
Qy 89 LeuThrLeuArgAlaGlyLeuAlaTyrAspLysAlaLeuSer----- 102  
Db 256 GTGATATTGCAGCAGTCCCTTAAGCTTGGTAAGCGCGATCAACAGCTGTAAACGGGAAT 315  
Qy 103 -----LysThrHisLeuSerAlaSerIleProAspThrAspArgMetTrpTyr 118  
Db 316 CTAATGATGCTAAAGAGCTAGTTGTTGCTAATGTTGGAGACTCCAGAGCAGTGTCT 375  
Qy 119 SerIleGlyAlaThrTyrLysPheThrProAsnLeuSerValAspValGlyPheAlaHis 138  
Db 376 AAGAACTGGTGTGGCAGTACAG-----CTATCTGTCTGAT-----CAT 411  
Qy 139 LeuArgGlyLysLysLysHisPheValGluThrGln-----AsnIleLys 153  
Db 412 GAACCAAGCAAGGAGAGAAAGAAATAGAGAGCGAGGTGGCTTTGTGTCAAATATTCCA 471  
Qy 154 Gly 154  
Db 472 CGG 474

RESULT 36  
US-09-974-300-1094  
;; Sequence 1094, Application US/09974300  
;; Patent No. US20020146721A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Berkas, Randy M.  
;; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
;; FILE REFERENCE: 10085.500-US  
;; CURRENT APPLICATION NUMBER: US/09/974,300  
;; PRIOR FILING DATE: 2001-10-05  
;; PRIOR APPLICATION NUMBER: 09/680,598  
;; PRIOR FILING DATE: 2000-10-06  
;; PRIOR APPLICATION NUMBER: 60/279,526

;; PRIOR FILING DATE: 2001-03-27  
;; NUMBER OF SEQ ID NOS: 8481  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 1094  
;; LENGTH: 636  
;; TYPE: DNA  
;; ORGANISM: Bacillus licheniformis  
US-09-974-300-1094

Alignment Scores: Length: 636  
Pred. No.: 3.05  
Score: 71.00  
Percent Similarity: 43.40%  
Best Local Similarity: 25.47%  
Query Match: 7.42%  
DB: 10

US-09-995-493-52 (1-179) x US-09-974-300-1094 (1-636)

Qy 54 GlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyrLysAsp 73  
Db 181 GGAAATTCAGTCAGTCGGGCTGCGGCTTTGAACAGCTTTCTGAACACGAAAGCTAAA--- 237  
Qy 74 AsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThrLeuArgAla 93  
Db 238 -----GTCGGCCGCTTTGAATGAGCTGAGCGAGCAATTT----- 270  
Qy 94 GlyLeuAlaTyrAspLysAlaLeuSerLysThrHisLeu---SerAlaSerIleProAsp 112  
Db 271 ---ATCCGGGAGACAAATAGCTGTTTGTACGCCGCTTTGGAAATTTCTCTTTCCCGCG 327  
Qy 113 ThrAspArgMetTrpTyrSer-----IleGlyAlaThrTyrLysPheThrPro 128  
Db 328 GTGATGAAGACGCTACTTCGATCGCTGCTAGCGGAAAGACGTTTAAATATACCGAA 387  
Qy 129 -----AsnLeuSerValAspValGlyPheAlaHisLeuArgGlyLysLys 144  
Db 388 CAGGACCGATCGGCTGTAAACGATAAAGACGCTGCACATCCAGCGAGAGCGGC 447  
Qy 145 HisPheValGluThrGln 150  
Db 448 TACTATTCAAGAGGCAG 465

RESULT 37

US-10-114-170-86  
;; Sequence 86, Application US/10114170  
;; Publication No. US20030023075A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Blattner, Frederick R.  
;; Burland, Valerie  
;; Perna, Nicole T.  
;; Plunkett, Guy  
;; Welch, Rod

TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Pinckney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53701-2113

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 Inch, 1.44MB storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 8.0  
CURRENT APPLICATION DATA:  
FILING DATE: 01-Apr-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION NUMBER: 09/453,702

; FILING DATE: 03-DEC-1999  
; APPLICATION NUMBER: 60/110,955  
; FILING DATE: 04-DEC-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J.  
; REGISTRATION NUMBER: 27386  
; REFERENCE/DOCKET NUMBER: 960296.95017  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (608) 251-5000  
; TELEFAX: (608) 251-9166  
; INFORMATION FOR SEQ ID NO: 86:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1548  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; SEQUENCE DESCRIPTION: SEQ ID NO: 86:  
US-10-114-170-86

Alignment Scores:  
Pred. No.: 11.2 Length: 1548  
Score: 71.00 Matches: 49  
Percent Similarity: 35.35% Conservativeness: 21  
Best Local Similarity: 24.75% Mismatches: 63  
Query Match: 7.42% Indels: 65  
DB: 9 Gaps: 10

US-09-995-493-52 (1-179) x US-10-114-170-86 (1-1548)

QY 15 LeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGlnLeu---ThrAsp 33  
DB 322 CTCAACTTATGTTTATGATGGCTTTGGTATCGGAGCGGTATACCAAAATCCGATCGGAC 381  
QY 34 GlnTrpAla----- 36  
DB 382 AAATCGCGAGGAAGAGCGCGTGAATCTCTCAATGCTCGGTAAAGTGCAGAACT 441  
QY 37 IleHisTyrSerTyr-LysTyrThrGlu-----TrpSerArgPheLysG1 51  
DB 442 GTGGCTACAGTATAAATATGATGCCAACACATTTTACTTTGCAGCTAATACGCTGA 501  
QY 51 uLeuArgGlyLysTyrGlnAspGlySerGlyTyr-----GluAlaPheTh 66  
DB 502 AACATTAACATGACACCATGATGGCGATGTTATATTTCTAAACAAAGCACAAAGTTTGA 561  
QY 66 rLysLysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThrTyrSerLeuAs 86  
DB 562 AGTGTGGCGCAATAT-----CAATTCGACTTCGGC----- 592  
QY 86 nAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLe 106  
DB 593 -----TTGGCGCCCTCACTCGCTACCTGAAA-----TCGAAAGGCATAGA 633  
QY 106 uSerAlaSerIleProAspThrAspArgMetTrpTyr---SerIleGlyAlaThrTyrLy 125  
DB 634 TCTGGCGCCCTACGGCGATCAGCATGATGATGATATATCGAGTGTGGTGGCAGCGATT 693  
QY 125 sPheThrProAsnLeuSerValAspValGlyPheAlaHisLeuArgGlyLysLysH1 145  
DB 694 CTTCAACAAAATATGTCGACCTATGTTGATTAT----- 727  
QY 145 sPheValGluThrGlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThrLys 165  
DB 728 -----AAATCAACCTGATTGATGAAGCGGACTTTACCCGTCG 765  
QY 165 sala-----ThrAlaAsnLeuTyrGlyLeuAsnLeuAsnTyrArgPhe 179  
DB 766 CGTAGATATTCGACCGCATAAATCGTCGTCACGGCGATTACCTATCATGTTTC 817  
RESULT 38  
US-09-822-849A-513  
; Sequence 513, Application US/09822849A

; Patent No. US20020045170A1  
; GENERAL INFORMATION:  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fechtel, Kim  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Howes, Steven H.  
; APPLICANT: Resnick, Richard J.  
; APPLICANT: Gulukota, Kamalakara  
; APPLICANT: Graham, James R.  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
; FILE REFERENCE: GIN 6403  
; CURRENT APPLICATION NUMBER: US/09/822,849A  
; PRIOR FILING DATE: 2001-09-04  
; PRIOR APPLICATION NUMBER: 60/195,582  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 598  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 513  
; LENGTH: 1365  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-822-849A-513  
Alignment Scores:  
Pred. No.: 10.9 Length: 1365  
Score: 70.50 Matches: 51  
Percent Similarity: 34.11% Conservativeness: 22  
Best Local Similarity: 23.83% Mismatches: 66  
Query Match: 7.37% Indels: 75  
DB: 10 Gaps: 11  
US-09-995-493-52 (1-179) x US-09-822-849A-513 (1-1365)  
QY 7 GlyProTyrIleGlyLysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSer 26  
DB 28 GGACCCACAGTGCATAAGGAACT-----CCACTGCCACTTACGAAGAGGCGCAAG 78  
QY 27 GlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrp 46  
DB 79 CAATAT-----CTGCTCTATGAAACGCTCTATGCCAATGGC 114  
QY 47 SerArgPheLysGluLeuArgGlyLysTyr----- 56  
DB 115 ACCCGCACAGACAGCGAGTGGGCATCTACATCCTCAGCAGTAGTGGAGTGGGCGCCAA 174  
QY 57 -----GlnAspGlySerGlyTyrGluAlaPheThrLysLysGluTyrLysAspAsn 74  
DB 175 CACCGAGACTCAGGGTCTTCAGGAAAGTCTCGAAGGAGCGGACGATTTATGGCTATGAC 234  
QY 75 SerArgPheAlaIle---GlyThrThrTyrSerLeuAsn-----AspAlaLeu 89  
DB 235 AGCAGGTTTCAGCATTTTGGGAAGGACTTCTCCGCTCAACATCCCTTTCACACATCAGTG 294  
QY 90 ThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSer 109  
DB 295 AAGTTATCCACGGCTGCACCGCGCACCCCTGGTGGCAGAGAAGCATCTCCTCAGCGTGCC 354  
QY 110 -----IleProAspThrAspArgMetTrpTyrSerIleGlyAlaThrTyr---LysPhe 126  
DB 355 CACTGCATACAGCAT-----GGAAAAACCTATGTGAAAGGA 390  
QY 127 ThrProAsnLeuSerVal----- 132  
DB 391 ACCCAGAGCTTCGAGTGGGCTTCCTTAAAGCCCAAGTTTAAAGATGGTGGTCCAGGGGCC 450  
QY 133 -----AspValGlyPheAlaHisLeuArgGlyLysLys 143  
DB 451 AACGACTCCACTTCAGCCATGCCCGCAGCAGATGAATTTTCAGTGGATCCGGTGAACGC 510  
QY 144 LysHisPheValGluThrGlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThr 163

```
Db 511 ACCCATGTGCC-----AAGGTTGGATC-----534
Qy 164 ThrLysAlaThrAlaAsnLeuTyrGlyLeuAsnLeuAsnTyr 177
Db 535 ---AAGGCAATGCCATGACATCGGATGATGATGATTAT 573

RESULT 39
US-09-738-626-216/c
; Sequence 216, Application US/09738526
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 216
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-216

Alignment Scores:
Pred. No.: 12.1 Length: 1464
Score: 70.50 Matches: 31
Percent Similarity: 38.10% Conservative: 25
Best Local Similarity: 21.09% Mismatches: 62
Query Match: 7.37% Indels: 29
DB: 9 Gaps: 6

US-09-995-493-52 (1-179) x US-09-738-626-216 (1-1464)
Qy 11 GlyLysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSerGlyPheHisGln 30
Db 951 GGAAGGGTCAGCTT---GTTCAAGTCCAGGATGGGAAGATATCTGGAACCTTCATCAG 895
Qy 31 LeuThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGlu---TrpSerArgPhe 49
Db 894 CTG-----TGGACATTAACCTTCAGTGGATCCTTGGCCAGCTTTGGTCCCAATC 844
Qy 50 LysGluLeu-----ArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPhe 65
Db 843 TTCAGCTGACCAAGTCTTCCAGGGAACCAACGAGATGTTGGAGCGGCAACGTTTGG 784
Qy 66 ThrLysLysGluGluTyrLysAspAsnSerArgPhe----- 77
Db 783 CTCAGGGCTGATGATCGAATCGAACCACGAGATAGTAGTGTGCTCTGATCGGTC 724
Qy 78 -----AlaIleGlyThrThrTyrSerLeuAsnAspAla 88
Db 723 GGAGTGAACGCAACGGTTTCTCCAAAGTTGTGGTCGGCTGCTCCCAATGAAGTA 664
Qy 89 LeuThrLeuArgAlaGlyLeuAla-----TyrAspLysAlaAlaSerLysThrHis 105
Db 663 GCGGTTTCCGTCCTTGGTCATCGGATGCGGTGACGACGATGATGCGGTTCAGGCCCAT 604
```

```
Qy 106 LeuSerAlaSerIleProAspThrAspArgMetTrpTyrSerIleGlyAlaThrTyrLys 125
Db 603 GCCACCAACGTTGCCCAACAGGTGCGCTGTGGTCTTCGCGGCTGCTTCTGCTC 544
Qy 126 PheThrProAsnLeuSerVal 132
Db 543 GAGGTGCAGGATCGGTCGCTC 523

RESULT 40
US-09-765-205-11
; Sequence 11, Application US/09765205
; Patent No.: US20020034800A1
; GENERAL INFORMATION:
; APPLICANT: CRO, LI
; TITLE OF INVENTION: BONE MARROW SECRETED PROTEINS AND POLYNUCLEOTIDES
; FILE REFERENCE: 1458.004/200130.449
; CURRENT APPLICATION NUMBER: US/09/765, 205
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US/09/212, 440
; PRIOR FILING DATE: 1998-12-16
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 1585
; TYPE: DNA
; ORGANISM: human
US-09-765-205-11

Alignment Scores:
Pred. No.: 13.6 Length: 1585
Score: 70.50 Matches: 51
Percent Similarity: 34.11% Conservative: 22
Best Local Similarity: 23.83% Mismatches: 66
Query Match: 7.37% Indels: 75
DB: 10 Gaps: 11

US-09-995-493-52 (1-179) x US-09-765-205-11 (1-1585)
Qy 7 GlyProTyrIleGlyLysGlySerLeuThrLeuLysLeuProAlaTyrTrpGluLeuSer 26
Db 236 GGACCCCGAGTGTCTAAGGAACT-----CCACTGCCACTTACGAAGAGGCCAAG 286
Qy 27 GlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrGluTrp 46
Db 287 CAATAT-----CTGTCTTTATGAAGGCTCTATGTCACATGGC 322
Qy 47 SerArgPheLysGluLeuArgGlyLysTyr----- 56
Db 323 AGCGGCACAGACGCGAGGTGGCATCTACATCTCAGCAGTACTGAGATGGGGCCAA 382
Qy 57 -----GlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyrLysAspAsn 74
Db 383 CACCGAGACTCAGGCTCTTCAGGAAGTCTCGAAGAAAGCGGAGATTTATGGCTATGAC 442
Qy 75 SerArgPheAlaIle---GlyThrThrTyrSerLeuAsn-----AspAlaLeu 89
Db 443 AGCAGGTTTCAGCATTTTGGGAAGGACTTCTGCTCAACTACCTTCTCAACATCAGTG 502
Qy 90 ThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSer 109
Db 503 AAGTTATCCAGGGCTGCACCGGACCCCTGGTGCAGAGAAGCATGCTCCACAGCTGCC 562
Qy 110 -----IleProAspThrAspArgMetTrpTyrSerIleGlyAlaThrTyr---LysPhe 126
Db 563 CACTGCATACAGCAT-----GGAAAAACCTATGTGAAAGGA 598
Qy 127 ThrProAsnLeuSerVal----- 132
Db 599 ACCCAGAAAGCTTCGAGTGGGCTTCCTAAAGCCCAAGTTTAAAGATGTGTGTCGAGGGGCC 658
Qy 133 -----AspValGlyPheAlaHisLeuArgGlyLysLys 143
Db 659 AACGACTCCACTTCAGCCATGCCCGGACGAGATGAATTTTCAGTGGATCCGGGTCAACGCG 718
```

Qy 144 LysHisPheValGluThrGlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThr 163  
 Db 719 ACCCATGTGCC-----AAGGTTGATC-----742

Qy 164 ThrLysAlaThrAlaAsnLeuTyrGlyLeuAsnLeuAsnTyr 177  
 Db 743 ---AAGGGCAATGCCAATGACATCGGCATGGATTATGATTAT 781

Search completed: May 19, 2003, 20:36:41  
 Job time : 174 secs





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 19, 2003, 19:03:23 ; Search time 1078 Seconds  
(without alignments)  
2689.232 Million cell updates/sec

Title: US-09-995-493-52  
Perfect score: 957  
Sequence: 1 OHNGVLGPIYGRKSLTLKP.....ADYTKATANLYGLNLNRYF 179

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-O=cn2.1/USPTO.spool/US0995493/runat\_12052003\_091309\_23015/app.query.fasta\_1.327  
-DB=EST -QFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi -LIST=45  
-DOCLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0995493@cncn\_1.1456 @runat\_12052003\_091309\_23015 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	304	31.8	997	17	CNS010FK	AL153456 Anopheles
C 2	120.5	12.6	796	17	BH385201	BH385201 AG-ND-160
C 3	100.5	10.5	746	17	BH383328	BH383328 AG-ND-160
C 4	96.5	10.1	659	13	BJ429628	-BJ429628 BJ429628
C 5	96.5	10.1	727	13	BJ429792	BJ429792 BJ429792
C 6	96.5	10.1	747	14	C91402	C91402 C91402 Dict
C 7	86	9.0	621	17	AQ990154	AQ990154 Rfc00876
C 8	85	8.9	941	14	BQ706806	BQ706806 AGENCOURT
C 9	84	8.8	672	13	B1547625	B1547625 G03191761
C 10	83.5	8.7	431	13	B1139361	B1139361 F130P15Y
C 11	83.5	8.7	756	12	BG064935	BG064935 H3025E07-
C 12	83	8.7	699	10	BE005948	BE005948 RCO-BN012
C 13	82.5	8.6	600	17	BH878604	BH878604 hs83b03 b
C 14	82.5	8.6	649	10	AW553447	AW553447 L0227C06-
C 15	82.5	8.6	683	13	BJ041818	BJ041818 BJ041818
C 16	82.5	8.6	744	14	BM930833	BM930833 UI-E-EJ1-
C 17	82	8.6	665	17	A2573577	A2573577 319PVA06
C 18	82	8.6	725	13	BM396839	BM396839 5009-0-25
C 19	81.5	8.5	386	10	AW788126	AW788126 C00212-F
C 20	81.5	8.5	417	10	AW759240	AW759240 sl38f12.y
C 21	81.5	8.5	563	13	BM178271	BM178271 saj71a04.y
C 22	81.5	8.5	1496	11	AK010501	AK010501 Mus muscu
C 23	81	8.5	659	12	BG740114	BG740114 G02631193
C 24	81	8.5	873	9	AL531424	AL531424 AL531424
C 25	80.5	8.4	484	14	W64749	W64749 md88c04.r1
C 26	80.5	8.4	484	14	W64824	W64824 md88g04.r1
C 27	80	8.4	516	14	BQ636341	BQ636341 hd07g02.y
C 28	80	8.4	571	9	AL701828	AL701828 DRFP0686H
C 29	80	8.4	598	13	B1819893	B1819893 G03035435
C 30	80	8.4	599	13	BM090644	BM090644 lg16g01.y
C 31	80	8.4	611	9	AL531052	AL531052 AL531052
C 32	80	8.4	615	12	BG700910	BG700910 G02681855
C 33	80	8.4	617	14	BQ639058	BQ639058 hd30h04.y
C 34	80	8.4	623	14	BQ639009	BQ639009 hd30c08.y
C 35	80	8.4	631	14	BQ639930	BQ639930 he22a07.y
C 36	80	8.4	647	14	BM707016	BM707016 UI-E-CQ0-
C 37	80	8.4	647	14	BQ636248	BQ636248 hd06g06.y
C 38	80	8.4	650	13	B1669859	B1669859 G03293455
C 39	80	8.4	662	14	BQ635968	BQ635968 hd02f11.y
C 40	80	8.4	665	10	BE006001	BE006001 RCO-BN012
C 41	80	8.4	707	12	BG829263	BG829263 G02753162
C 42	80	8.4	715	12	BG705579	BG705579 G02668762
C 43	80	8.4	739	12	BG708348	BG708348 G02672287
C 44	80	8.4	780	13	B1549564	B1549564 G03192261
C 45	80	8.4	791	13	BM399133	BM399133 5009-0-53

ALIGNMENTS

RESULT 1  
CNS010FK/c  
LOCUS  
DEFINITION  
CNS010FK  
Anopheles gambiae GSS SP6 end of clone 26E01 of Notredame library  
from strain PEST of Anopheles gambiae (African malaria mosquito),  
genomic survey sequence.  
AL153456  
AL153456.1 GI:7014375  
GSS.  
SOURCE  
African malaria mosquito.  
ORGANISM  
Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
Anopheles.  
REFERENCE  
1 (bases 1 to 997)

**AUTHORS** Genoscope.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [segrete@genoscope.cns.fr](mailto:segrete@genoscope.cns.fr)  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))

DEFINITION	AG-ND-160D9.TR ND-TAM Anopheles gambiae genomic clone AG-ND-160D9, DNA sequence.
ACCESSION	BH385201
VERSION	BH385201.1
KEYWORDS	GI:17331343
SOURCE	GSS:
ORGANISM	African malaria mosquito. Anopheles gambiae Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
REFERENCE	1 (bases 1 to 796)
AUTHORS	Shetty, V., Malek, O., Koo, H., Collins, F., Gardner, M. and Loftus, B. J.
TITLE	Direct Substitution of BAC-end sequences from Anopheles gambiae
JOURNAL	Unpublished (2001).
COMMENT	Other GSSs: AG-ND-160D9.TF

REFERENCE  
1 (bases 1 to 796)  
AUTHORS  
Shetty J., Matek J., Koo H., Collins F., Gardner M. and Loftus B.J.  
TITLE  
Direct Submission of BAC-end sequences from Anopheles gambiae  
JOURNAL  
Unpublished (2001)  
COMMENT  
Other GSST: AG-ND-160D9.TF  
neoptera, emphytreroform, diptera, nematocera, culicidae,  
Anopheles.

Tel.: 301 838 3503  
 Fax: 301 838 3543  
 Email: bjloftus@tigr.org  
 This clone is from an *A. gambiae* BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from *A. gambiae* PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit

library was constructed at Texas A&M University BAC Center

University, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
Seq primer: M13 Rev  
Class: BAC ends.

```

FEATURES
  source
    Location/Qualifiers
      1..796
        /organism="Anopheles gambiae"
        /strain="PEST"

```

```

return bsl
/db_xref=taxon:7165"
/clone="AG-ND-160D9"
/clone_lib="ND-TAM"

```

```

BASE COUNT
ORIGIN
232 a 142 c 154 g 268 t
/notice=vector; Site_1: HindIII*

```

ORIGIN  
Alignment Scores:  
pred. No.: 8.37e-05 Length: 796

Pred. NO.:	0.37E+05	750
Score:	120.50	33
Percent Similarity:	39.88%	34
Best Local Similarity:	19.64%	60
Mismatches:		
Matches:		
Length:		

```
DB: 17
Query Match: 17
Best Local Similarity: 17.04%
Query Match: 17.59%
Indels: 41
Mismatches: 80
Gaps: 5
```

US-09-995-493-52 (1-179) x BH385201 (1-796)

QY 30 GlnLeu<sup>30</sup>AspGlnIrrp<sup>31</sup>AlaIleu<sup>32</sup>His<sup>33</sup>Leu<sup>34</sup>Ser<sup>35</sup>Leu<sup>36</sup>Val<sup>37</sup>Trp<sup>38</sup>Arg<sup>39</sup>Pro<sup>40</sup>Arg<sup>41</sup>Pro<sup>42</sup>Arg<sup>43</sup>Pro<sup>44</sup>Arg<sup>45</sup>Pro<sup>46</sup>Arg<sup>47</sup>Pro<sup>48</sup>Arg<sup>49</sup>Pro<sup>50</sup>Arg<sup>51</sup>Pro<sup>52</sup>Arg<sup>53</sup>Pro<sup>54</sup>Arg<sup>55</sup>Pro<sup>56</sup>Arg<sup>57</sup>Pro<sup>58</sup>Arg<sup>59</sup>Pro<sup>60</sup>Arg<sup>61</sup>Pro<sup>62</sup>Arg<sup>63</sup>Pro<sup>64</sup>Arg<sup>65</sup>Pro<sup>66</sup>Arg<sup>67</sup>Pro<sup>68</sup>Arg<sup>69</sup>Pro<sup>70</sup>Arg<sup>71</sup>Pro<sup>72</sup>Arg<sup>73</sup>Pro<sup>74</sup>Arg<sup>75</sup>Pro<sup>76</sup>Arg<sup>77</sup>Pro<sup>78</sup>Arg<sup>79</sup>Pro<sup>80</sup>Arg<sup>81</sup>Pro<sup>82</sup>Arg<sup>83</sup>Pro<sup>84</sup>Arg<sup>85</sup>Pro<sup>86</sup>Arg<sup>87</sup>Pro<sup>88</sup>Arg<sup>89</sup>Pro<sup>90</sup>Arg<sup>91</sup>Pro<sup>92</sup>Arg<sup>93</sup>Pro<sup>94</sup>Arg<sup>95</sup>Pro<sup>96</sup>Arg<sup>97</sup>Pro<sup>98</sup>Arg<sup>99</sup>Pro<sup>100</sup>Arg<sup>101</sup>Pro<sup>102</sup>Arg<sup>103</sup>Pro<sup>104</sup>Arg<sup>105</sup>Pro<sup>106</sup>Arg<sup>107</sup>Pro<sup>108</sup>Arg<sup>109</sup>Pro<sup>110</sup>Arg<sup>111</sup>Pro<sup>112</sup>Arg<sup>113</sup>Pro<sup>114</sup>Arg<sup>115</sup>Pro<sup>116</sup>Arg<sup>117</sup>Pro<sup>118</sup>Arg<sup>119</sup>Pro<sup>120</sup>Arg<sup>121</sup>Pro<sup>122</sup>Arg<sup>123</sup>Pro<sup>124</sup>Arg<sup>125</sup>Pro<sup>126</sup>Arg<sup>127</sup>Pro<sup>128</sup>Arg<sup>129</sup>Pro<sup>130</sup>Arg<sup>131</sup>Pro<sup>132</sup>Arg<sup>133</sup>Pro<sup>134</sup>Arg<sup>135</sup>Pro<sup>136</sup>Arg<sup>137</sup>Pro<sup>138</sup>Arg<sup>139</sup>Pro<sup>140</sup>Arg<sup>141</sup>Pro<sup>142</sup>Arg<sup>143</sup>Pro<sup>144</sup>Arg<sup>145</sup>Pro<sup>146</sup>Arg<sup>147</sup>Pro<sup>148</sup>Arg<sup>149</sup>Pro<sup>150</sup>Arg<sup>151</sup>Pro<sup>152</sup>Arg<sup>153</sup>Pro<sup>154</sup>Arg<sup>155</sup>Pro<sup>156</sup>Arg<sup>157</sup>Pro<sup>158</sup>Arg<sup>159</sup>Pro<sup>160</sup>Arg<sup>161</sup>Pro<sup>162</sup>Arg<sup>163</sup>Pro<sup>164</sup>Arg<sup>165</sup>Pro<sup>166</sup>Arg<sup>167</sup>Pro<sup>168</sup>Arg<sup>169</sup>Pro<sup>170</sup>Arg<sup>171</sup>Pro<sup>172</sup>Arg<sup>173</sup>Pro<sup>174</sup>Arg<sup>175</sup>Pro<sup>176</sup>Arg<sup>177</sup>Pro<sup>178</sup>Arg<sup>179</sup>Pro<sup>180</sup>Arg<sup>181</sup>Pro<sup>182</sup>Arg<sup>183</sup>Pro<sup>184</sup>Arg<sup>185</sup>Pro<sup>186</sup>Arg<sup>187</sup>Pro<sup>188</sup>Arg<sup>189</sup>Pro<sup>190</sup>Arg<sup>191</sup>Pro<sup>192</sup>Arg<sup>193</sup>Pro<sup>194</sup>Arg<sup>195</sup>Pro<sup>196</sup>Arg<sup>197</sup>Pro<sup>198</sup>Arg<sup>199</sup>Pro<sup>200</sup>Arg<sup>201</sup>Pro<sup>202</sup>Arg<sup>203</sup>Pro<sup>204</sup>Arg<sup>205</sup>Pro<sup>206</sup>Arg<sup>207</sup>Pro<sup>208</sup>Arg<sup>209</sup>Pro<sup>210</sup>Arg<sup>211</sup>Pro<sup>212</sup>Arg<sup>213</sup>Pro<sup>214</sup>Arg<sup>215</sup>Pro<sup>216</sup>Arg<sup>217</sup>Pro<sup>218</sup>Arg<sup>219</sup>Pro<sup>220</sup>Arg<sup>221</sup>Pro<sup>222</sup>Arg<sup>223</sup>Pro<sup>224</sup>Arg<sup>225</sup>Pro<sup>226</sup>Arg<sup>227</sup>Pro<sup>228</sup>Arg<sup>229</sup>Pro<sup>230</sup>Arg<sup>231</sup>Pro<sup>232</sup>Arg<sup>233</sup>Pro<sup>234</sup>Arg<sup>235</sup>Pro<sup>236</sup>Arg<sup>237</sup>Pro<sup>238</sup>Arg<sup>239</sup>Pro<sup>240</sup>Arg<sup>241</sup>Pro<sup>242</sup>Arg<sup>243</sup>Pro<sup>244</sup>Arg<sup>245</sup>Pro<sup>246</sup>Arg<sup>247</sup>Pro<sup>248</sup>Arg<sup>249</sup>Pro<sup>250</sup>Arg<sup>251</sup>Pro<sup>252</sup>Arg<sup>253</sup>Pro<sup>254</sup>Arg<sup>255</sup>Pro<sup>256</sup>Arg<sup>257</sup>Pro<sup>258</sup>Arg<sup>259</sup>Pro<sup>260</sup>Arg<sup>261</sup>Pro<sup>262</sup>Arg<sup>263</sup>Pro<sup>264</sup>Arg<sup>265</sup>Pro<sup>266</sup>Arg<sup>267</sup>Pro<sup>268</sup>Arg<sup>269</sup>Pro<sup>270</sup>Arg<sup>271</sup>Pro<sup>272</sup>Arg<sup>273</sup>Pro<sup>274</sup>Arg<sup>275</sup>Pro<sup>276</sup>Arg<sup>277</sup>Pro<sup>278</sup>Arg<sup>279</sup>Pro<sup>280</sup>Arg<sup>281</sup>Pro<sup>282</sup>Arg<sup>283</sup>Pro<sup>284</sup>Arg<sup>285</sup>Pro<sup>286</sup>Arg<sup>287</sup>Pro<sup>288</sup>Arg<sup>289</sup>Pro<sup>290</sup>Arg<sup>291</sup>Pro<sup>292</sup>Arg<sup>293</sup>Pro<sup>294</sup>Arg<sup>295</sup>Pro<sup>296</sup>Arg<sup>297</sup>Pro<sup>298</sup>Arg<sup>299</sup>Pro<sup>300</sup>Arg<sup>301</sup>Pro<sup>302</sup>Arg<sup>303</sup>Pro<sup>304</sup>Arg<sup>305</sup>Pro<sup>306</sup>Arg<sup>307</sup>Pro<sup>308</sup>Arg<sup>309</sup>Pro<sup>310</sup>Arg<sup>311</sup>Pro<sup>312</sup>Arg<sup>313</sup>Pro<sup>314</sup>Arg<sup>315</sup>Pro<sup>316</sup>Arg<sup>317</sup>Pro<sup>318</sup>Arg<sup>319</sup>Pro<sup>320</sup>Arg<sup>321</sup>Pro<sup>322</sup>Arg<sup>323</sup>Pro<sup>324</sup>Arg<sup>325</sup>Pro<sup>326</sup>Arg<sup>327</sup>Pro<sup>328</sup>Arg<sup>329</sup>Pro<sup>330</sup>Arg<sup>331</sup>Pro<sup>332</sup>Arg<sup>333</sup>Pro<sup>334</sup>Arg<sup>335</sup>Pro<sup>336</sup>Arg<sup>337</sup>Pro<sup>338</sup>Arg<sup>339</sup>Pro<sup>340</sup>Arg<sup>341</sup>Pro<sup>342</sup>Arg<sup>343</sup>Pro<sup>344</sup>Arg<sup>345</sup>Pro<sup>346</sup>Arg<sup>347</sup>Pro<sup>348</sup>Arg<sup>349</sup>Pro<sup>350</sup>Arg<sup>351</sup>Pro<sup>352</sup>Arg<sup>353</sup>Pro<sup>354</sup>Arg<sup>355</sup>Pro<sup>356</sup>Arg<sup>357</sup>Pro<sup>358</sup>Arg<sup>359</sup>Pro<sup>360</sup>Arg<sup>361</sup>Pro<sup>362</sup>Arg<sup>363</sup>Pro<sup>364</sup>Arg<sup>365</sup>Pro<sup>366</sup>Arg<sup>367</sup>Pro<sup>368</sup>Arg<sup>369</sup>Pro<sup>370</sup>Arg<sup>371</sup>Pro<sup>372</sup>Arg<sup>373</sup>Pro<sup>374</sup>Arg<sup>375</sup>Pro<sup>376</sup>Arg<sup>377</sup>Pro<sup>378</sup>Arg<sup>379</sup>Pro<sup>380</sup>Arg<sup>381</sup>Pro<sup>382</sup>Arg<sup>383</sup>Pro<sup>384</sup>Arg<sup>385</sup>Pro<sup>386</sup>Arg<sup>387</sup>Pro<sup>388</sup>Arg<sup>389</sup>Pro<sup>390</sup>Arg<sup>391</sup>Pro<sup>392</sup>Arg<sup>393</sup>Pro<sup>394</sup>Arg<sup>395</sup>Pro<sup>396</sup>Arg<sup>397</sup>Pro<sup>398</sup>Arg<sup>399</sup>Pro<sup>400</sup>Arg<sup>401</sup>Pro<sup>402</sup>Arg<sup>403</sup>Pro<sup>404</sup>Arg<sup>405</sup>Pro<sup>406</sup>Arg<sup>407</sup>Pro<sup>408</sup>Arg<sup>409</sup>Pro<sup>410</sup>Arg<sup>411</sup>Pro<sup>412</sup>Arg<sup>413</sup>Pro<sup>414</sup>Arg<sup>415</sup>Pro<sup>416</sup>Arg<sup>417</sup>Pro<sup>418</sup>Arg<sup>419</sup>Pro<sup>420</sup>Arg<sup>421</sup>Pro<sup>422</sup>Arg<sup>423</sup>Pro<sup>424</sup>Arg<sup>425</sup>Pro<sup>426</sup>Arg<sup>427</sup>Pro<sup>428</sup>Arg<sup>429</sup>Pro<sup>430</sup>Arg<sup>431</sup>Pro<sup>432</sup>Arg<sup>433</sup>Pro<sup>434</sup>Arg<sup>435</sup>Pro<sup>436</sup>Arg<sup>437</sup>Pro<sup>438</sup>Arg<sup>439</sup>Pro<sup>440</sup>Arg<sup>441</sup>Pro<sup>442</sup>Arg<sup>443</sup>Pro

Qy 50 LysGluLeuArgGLylstyr-----GlnAspGlySerGlyTyrGlu 63  
          :::|::|               :::  
Db 632 AGCAACCTGAACATTAACTTGCTTAACGCCAAAATAGGAATGACCACCGCATTCAGA 573

DB	632	AGCAAACTGAAACATTACCTTTGCTTAACGCCAAATATTAGGAAATAGACCAACCGATATGAGA	573
Qy	64	AlaPheThrLysLysGluGluTyrlLysAspAsnSerArgPheAlaIleGlyThrThrTrl	83

Db 572 GTTTATTTCAGCTCCTCTAAAAAATTCACAAAACACAAATCTTATAGACTCGGAACAATAC 513  
::: ::::: :  
Ov 84 SerLeuAsnAspAlaIeuThrLeuArgAlaClcVleAlaLysTrAspLysAlaAla--Ser 102

```

Db 512 TGGTTTACTGATATGATGCCGGAAGACTTGGATATTACTATGACGAGTCTCCTTACAAT 453
Qy 103 LysThrHisLeuSerAlaSerIleProAspThrAsp----- 114
Db 452 GATAACCAATTAGCCCGAAACTCCTTCTATGACACACAATGTACTACAGGAGGTATC 393
Qy 115 -----ArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsn 129
Db 392 GGTCTTAAGTTTGGAAAACTTGGAGTAGACTTTAGCAGCGCAGCTACAATTTCTACAAGAGC 333
Qy 130 LeuSerValAspValGlyPheAlaHisLeuArgGlyLysLysLysHisPheValGluThr 149
Db 332 AGAAGTGTAGCAATGACTTCTACACCAATGACGCGACAG----- 294
Qy 150 GlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThrLysAlaThrAlaAsn 169
Db 293 -----ATAAAGGA-----AGTCTTTT 276

Qy 170 LeuTyrGlyLeuLeuAsnTyr 177
Db 275 ATCTTCGGACTTGGCCTTTTCATAT 252

RESULT 3
BH383328/c
LOCUS BH383328 746 bp DNA linear GSS 10-DEC-2001
DEFINITION AG-ND-160D14.TR ND-TAM Anopheles gambiae genomic clone AG-ND-160D14
, DNA sequence.
ACCESSION BH383328
VERSION BH383328.1 GI:17329470
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
1 (bases 1 to 746)
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
Direct Submission of BAC-end sequences from Anopheles gambiae
Unpublished (2001)
Other_GSSs: AG-ND-160D14.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@igr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 Rev
Class: BAC ends.
Location/Qualifiers
1..746
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-160D14"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 219 a 132 c 141 g 254 t
ORIGIN

Alignment Scores:
Pred. No.: 0.0288 Length: 746
Score: 100.50 Matches: 32
Percent Similarity: 38.79% Conservative: 32

```

```

Best Local Similarity: 19.39% Mismatches: 59
Query Match: 10.50% Indels: 42
DB: 17 Gaps: 5

US-09-995-493-52 (1-179) x BH383328 (1-746)

Qy 34 GlnTTPAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGluLeuArg 53
Db 681 AGATGCTCAGTTTCTGGGGATTTCAACTATTACGGATGGGATCGTTTACAGCAACCTGACC 622
Qy 54 GlyLysTyr-----GlnAspGlySerGlyTyrCluAlaPheThrLys 67
Db 621 ATTAATTTGCTTAACGCAAAATTAGAAATGACCCAAAGCATATGAGAGTTTATTACAGCT 562
Qy 68 LysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThr-ThrTyrSerLeuAsnAs 87
Db 561 CCTAAAACTTCAAAAACACAAAATCTTATAGAGTCGGAACACAAATATGTTTACTGA 502
Qy 87 pAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAla---SerLysThrHisLe 106
Db 501 TATGATTGCCGGAAGACTTGGATATTACTATGACGAGTCTCCTTACAATGATAACCAATT 442
Qy 106 uSerAlaSerIleProAspThrAsp----- 114
Db 441 TAGCCGGAAATCTCCTTCTATGAACAACAATGTACTAACAGGAGGTATCGGTTTAAGTT 382
Qy 115 ----ArgMetTrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSerValAs 133
Db 381 TGGAAACCTTGGAGTAGACTTAGCAGCCAGCTACAAATTTCTACAGAGCAGAACTGTTAG 322
Qy 133 pValGlyPheAlaHisLeuArgGlyLysLysHisPheValGluThrGlnAsnIleLys 153
Db 321 CAATGACTTCTACACATGAGCGGACAG-----ATAA 289
Qy 153 sGlyLeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeuTyrClyLe 173
Db 288 AGGA-----AGTGCTTTTATCTTCGGACT 265
Qy 173 uAsnLeuAsnTyr 177
Db 264 TGGCCTTTTCATAT 252

RESULT 4
BH429628/c
LOCUS BH429628 659 bp mRNA linear EST 13-MAR-2002
DEFINITION BH429628 Dictyostelium discoideum cDNA library, VF Dictyostelium
discoideum cDNA clone ddv4d10 3', mRNA sequence.
ACCESSION BH429628.1 GI:19404350
VERSION BH429628
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
REFERENCE 1 (bases 1 to 659)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
TITLE Full length cDNA of Dictyostelium discoideum at the vegetative
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Location/Qualifiers
1..659
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddv4d10"
/clone_lib="Dictyostelium discoideum cDNA library, VF"
/sex="mat A"

FEATURES
source

```

BASE COUNT 231 a 118 c 90 g 220 t  
ORIGIN

Alignment Scores: /dev\_stage="growth phase"  
Pred. No.: 0.0781 Length: 659  
Score: 96.50 Matches: 39  
Percent Similarity: 42.47% Conservative: 23  
Best Local Similarity: 26.71% Mismatches: 53  
Query Match: 10.08% Indels: 31  
DB: 13 Gaps: 7

US-09-995-493-52 (1-179) x BJ429628 (1-659)

QY 38 HistTyrSerTyrLysTyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGln 57  
|||||  
Db 418 CATTATTCGTAT-----AGTAAATATCGT 395

QY 58 AspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyrLysAspAsnSerArgPhe 77  
|||||  
Db 394 AATGGAATTCATATTTGGCATTGGTTTAAAGAAGCTTTTAGAATGTATCCAGCTGGT 335

QY 78 AlailegylThrThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyr 97  
|||||  
Db 334 ATTTAGTTTACCTCATATGACTTCTGAGGATTGTCAGATTCAAGCT-----287

QY 98 AspLysAlaAlaSerLysThrHisLeuSerAlaSerIleProAspThrAspArgMetTrp 117  
|||||  
Db 286 CATAAATTCGAAGGTACTCAATCATTCANAAATATT-----248

QY 118 TyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSerValAspValGlyPheAla 137  
|||||  
Db 247 TACTCTACTCATGCTAGTGAATCTTTTGGCCAAATCCAAATATTTATACACCAAGAAGA 188

QY 138 HisLeuArg-----GlyLysLysLysHisPhe---ValGluThrGlnAsnIleLys 153  
|||||  
Db 187 CATATTCAAATGATGTTCTAAAGTGTTCATTTCCCTGTTGGTACTAGAAATTTGATG 128

QY 154 GlyLeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeuTyrGlyLeu 173  
|||||  
Db 127 GGTATGAGTTTAAAGTGGCGTGAAGTTCATACT---GCTATGGCTGAATTTATTTGGT---74

QY 174 AsnLeuAsnTyrArgPhe 179  
|||||  
Db 73 -----AATTTCAAATTT 62

RESULT 5  
BJ429792/c 727 bp mRNA linear EST 13-MAR-2002  
LOCUS  
DEFINITION BJ429792 Dictyostellium discoideum cDNA library, VF Dictyostellium  
discoideum cDNA clone ddv4k24 3', mRNA sequence.

ACCESSION  
VERSION BJ429792  
KEYWORDS  
SOURCE EST.  
ORGANISM Dictyostellium discoideum.

REFERENCE  
AUTHORS Eukaryota; Mycetozoa; Dictyostelid; Dictyostellium.  
1 (bases 1 to 727)

TITLE  
JOURNAL Full length cDNA of Dictyostellium discoideum at the vegetative  
stage

COMMENT  
Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp.  
Location/Qualifiers

FEATURES  
source  
1. .727  
/organism="Dictyostellium discoideum"  
/strain="AX4"

/db\_xref="taxon:44689"  
/clone="ddv4k24"  
/clone\_lib="Dictyostellium discoideum cDNA library, VF"  
/sex="mat A"  
/dev\_stage="growth phase"  
BASE COUNT 245 a 124 c 98 g 256 t 4 others  
ORIGIN

Alignment Scores: Length: 727  
Pred. No.: 0.0909 Matches: 39  
Score: 96.50 Conservative: 23  
Percent Similarity: 42.47% Mismatches: 53  
Best Local Similarity: 26.71% Indels: 31  
Query Match: 10.08% Gaps: 7  
DB: 13

US-09-995-493-52 (1-179) x BJ429792 (1-727)

QY 38 HistTyrSerTyrLysTyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGln 57  
|||||  
Db 460 CATTATTCGTAT-----AGTAAATATCGT 437

QY 58 AspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyrLysAspAsnSerArgPhe 77  
|||||  
Db 436 AATGGAATTCATATTTGGCATTGGTTTAAAGAAGCTTTTAGAATGTATCCAGCTGGT 377

QY 78 AlailegylThrThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyr 97  
|||||  
Db 376 ATTTAGTTTACCTCATATGACTTCTGAGGATTGTCAGATTCAAGCT-----329

QY 98 AspLysAlaAlaSerLysThrHisLeuSerAlaSerIleProAspThrAspArgMetTrp 117  
|||||  
Db 328 CATAAATTCGAAGGTACTCAATCATTCANAAATATT-----290

QY 118 TyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSerValAspValGlyPheAla 137  
|||||  
Db 289 TACTCTACTCATGCTAGTGAATCTTTTGGCCAAATCCAAATATTTATACACCAAGAAGA 230

QY 138 HisLeuArg-----GlyLysLysLysHisPhe---ValGluThrGlnAsnIleLys 153  
|||||  
Db 229 CATATTCAAATGATGTTCTAAAGTGTTCATTTCCCTGTTGGTACTAGAAATTTGATG 170

QY 154 GlyLeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeuTyrGlyLeu 173  
|||||  
Db 169 GGTATGAGTTTAAAGTGGCGTGAAGTTCATACT---GCTATGGCTGAATTTATTTGGT---116

QY 174 AsnLeuAsnTyrArgPhe 179  
|||||  
Db 115 -----AATTTCAAATTT 104

RESULT 6  
C91402

LOCUS  
DEFINITION C91402 Dictyostellium discoideum SS (H. Urushihara) Dictyostellium  
discoideum cDNA clone SSK171, mRNA sequence.

ACCESSION  
VERSION C91402.1 GI:3060768  
KEYWORDS  
SOURCE EST.

ORGANISM Dictyostellium discoideum.  
Dictyostellium discoideum  
Dictyostellium discoideum  
Eukaryota; Mycetozoa; Dictyostelid; Dictyostellium.

REFERENCE  
AUTHORS Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,  
Yoshino, R., Mitra, B. N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,  
Williams, J., Maeda, M., Takeuchi, I., Ochial, H. and Tanaka, Y.  
Developmental cDNA in Dictyostellium discoideum  
Unpublished (1998)

JOURNAL  
COMMENT Contact: Hideko Urushihara  
Institute of Biological Sciences  
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan  
Tel: 81-298-53-4664  
Fax: 81-298-53-6614

Email: hideko@blol.tsukuba.ac.jp.  
 Location/Qualifiers  
 1. 747  
 /organism="Dictyostellium discoideum"  
 /strain="AX4"  
 /db\_xref="taxon:44689"  
 /clone="SSK171"  
 /dev\_stage="slug"  
 298 a 81 c 116 g 252 t  
 BASE COUNT  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.0948 Length: 747  
 Score: 96.50 Matches: 39  
 Percent Similarity: 42.47% Conservative: 23  
 Best Local Similarity: 26.71% Mismatches: 53  
 Query Match: 10.08% Indels: 31  
 DB: 14 Gaps: 7

US-09-995-493-52 (1-179) x C91402 (1-747)  
 QY 38 HisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGln 57  
 Db 202 CATATTCGTAT-----AGTAAATATCGT 225  
 QY 58 AspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyrLysAspAsnSerArgPhe 77  
 Db 226 ATGGAAATCCATATTTGGCATGGTGTAAAGAAAGTTTATAGATGATCCAGCTGGT 285  
 QY 78 AlaIleGlyThrTyrThrSerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyr 97  
 Db 286 ATTTAGGTTTACCTCATATGACTTCTGAGATTGTGAGATTCAAGGT----- 333  
 QY 98 AspLysAlaAlaSerLysThrHisLeuSerAlaSerIleProaspThrAspArgMetTrp 117  
 Db 334 CATAAATTCGAAAGGGTACTCAATTCATTCATAAATATT----- 372  
 QY 118 TyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSerValAspValGlyPheAla 137  
 Db 373 TACTCTACTCATCGTAGTAAGTCTTTTGGCCAAATCCAAATTAATTCATACCAGAAAGA 432  
 QY 138 HisLeuArg-----GlyLysLysLysHisPhe---ValGluThrGlnAsnIleLys 153  
 Db 433 CATATCAAAATGATCTTTCTAAAGCTGTTTATTCGCTGTTGGTACTAGAAATTCATG 492  
 QY 154 GlyLeuLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeuTyrGlyLeu 173  
 Db 493 GGTATGAGTTTAAAGTGAAGCTCAAGTTCATACT---GCTATGCTCAATTTATTTGCT--- 546  
 QY 174 AsnLeuAsnTyrArgPhe 179  
 Db 547 -----AATTCAAATTT 558

RESULT 7  
 AQ990154/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Photobacterium.  
 1 (bases 1 to 621)  
 french-Constant R.H., Waterfield N., Burland V., Perna N.T.,  
 Daborn P.J., Bowen D. and Blattner F.R.  
 A genomic sample sequence of the entomopathogenic bacterium  
 Photobacterium luminescens W14: potential implications for virulence  
 Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)  
 20378633

Contact: french-Constant RH  
 Department of Biology and Biochemistry  
 University of Bath  
 South Building, Bath BA2 7AY, UK  
 Tel: (44) 1225 826621  
 Fax: (44) 1225 826779  
 Email: bssrfc@bath.ac.uk  
 This is one of 2,122 random reads from the M13 library. For  
 annotation of identified clones (BLASTX, BLASTN and mapping to E.  
 coli K12 genome) please see french-Constant et al. 2000, Nucleic  
 Acids Res.  
 Seq primer: M13 Forward  
 Class: shotgun.  
 Location/Qualifiers  
 1. 621  
 /organism="Photobacterium luminescens"  
 /strain="W14"  
 /db\_xref="taxon:29488"  
 /clone="PLG00876"  
 /clone\_lib="Photobacterium luminescens strain W14 M13  
 library"  
 /dev\_stage="primary phase variant"  
 /note="genomic DNA from strain W14 was size selected (1-2  
 kb) and then cloned into M13 Janus."  
 BASE COUNT 183 a 111 c 109 g 216 t 2 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.61 Length: 621  
 Score: 86.00 Matches: 38  
 Percent Similarity: 37.80% Conservative: 24  
 Best Local Similarity: 23.17% Mismatches: 59  
 Query Match: 8.99% Indels: 43  
 DB: 17 Gaps: 7

US-09-995-493-52 (1-179) x AQ990154 (1-621)  
 QY 29 HisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArg 48  
 Db 539 TATGACCTTGATGATCAATGGGGCTATATCATCATCATTTACNTTGTCT--- 492  
 QY 49 PheLys-GluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLys 68  
 Db 491 ---AAAGGAGATGAAAAGAACCAACCAAGGAATCCG-----ACCTCAA 450  
 QY 68 sGluGluTyrLysAspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAl 88  
 Db 449 AGTGATATTAAGTATTTCCCTAATGGCG---GGCCCAACCTATAGGATTATGATTA 393  
 QY 88 aLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLys-----ThrHisLe 105  
 Db 392 TATTAGCTTATATGGCCAACTGGGATTTATCTAGGATTAAATACAAATTTTCCACTCATT 333  
 QY 106 uSerAlaSerIleProaspThrAspArg-----MetTrpTyrSerIleG1 121  
 Db 332 TAGTACTGGAGATACCGAAAGGATGAAGCAAGCAAAATACCTTTAGTTGGGAGCGG 273  
 QY 121 yAlaThrTyrLysPheThrProAsnLeuSerValAspValGlyPheAlaHisLeuArg-- 140  
 Db 272 TTTTATTATTAATCCAAACCACTTCTCAATTTACCGCGGTATGAAAGGTAGCCGCTT 213  
 QY 141 -----GlyLysLysLysHisPheValGluThrGlnAsnIleLysGlyLeuLe 156  
 Db 212 TAGTATTAAAGACGGAACCAAGAGATCATTTATCTACT----- 173  
 QY 156 uLeuValGluAlaAspTyrThrThrLysAlaThrAlaAsnLeuTyrGlyLeuAsnLeuAs 176  
 Db 172 -----AACGGTTTAAATATACTGTCGG 150  
 QY 176 nTyrArgPhe 179  
 Db 149 CTATCGTTTC 140

```

RESULT 8
BQ706806      941 bp      mRNA      linear      EST 16-JUL-2002
LOCUS      AGENCOURT_8418322 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6281413
DEFINITION      5', mRNA sequence.
ACCESSION      BQ706806
VERSION      BQ706806.1 GI:21845705
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgapbs-femail.nih.gov
Tissue Procurement: Dr. Mark Watson
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2474 row: c column: 14
High quality sequence stop: 621.
FEATURES      source
Location/Qualifiers
1..941
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6281413"
/lab_host="DH10B (phage-resistant)"
/note="Organ: spleen; Vector: pORB7; Site1: XhoI; Site2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC library."
BASE COUNT      219 a      220 c      281 g      220 t      1 others
ORIGIN
Alignment Scores:
Pred. No.:      4.13      Length:      941
Score:      85.00      Matches:      42
Percent Similarity:      38.73%      Conservative:      25
Best Local Similarity:      24.28%      Mismatches:      65
Query Match:      8.88%      Indels:      43
DB:      14      Gaps:      8

US-09-995-493-52 (1-179) x BQ706806 (1-941)
QY      3      AsnGlyValLeuGlyProTyrIleGlyGlySerLeuThrLeuLysLeuProAlaTyr      22
Db      382      AATGGGGAGCTGTCCTATGTGGGCAAGCTGGC---ACTGTACCTTCCT-----      432
QY      23      TrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyr-----      39
Db      433      -----AGGAACCTCCCACTACCCCTCATGTTATAGTCGCCACTATGAAGATGGT      483
QY      40      -----SeryTyrLysTyrThrGluTrpSerArgPheLysGluLeuArgGlyLys      55
Db      484      TATCCAGGTGGCAGTATACTATGGCAGTCTGTCGGGGTGACCCGCGATTGAGGACGG      543
QY      56      TyrGlnAspGly----SerGlyTyrGluAlaPheThrLysLysGluGluTyrLysAspAsn      74
Db      544      TATAGGCCAGCATGGAAGGCTACCGGCACCTAGTAGACAGGATGTGTATGGGCCCAAA      603
QY      75      SerArgPheAlaIleGlyThrThr-----Tyr      83
Db      :      :      :      :      :      :      :      :      :      :

```

```

Db      604      CCCAGGTCGGTAGTGGGAGCAGCGTGGATCTGCATCGCTTTCATCCAGACGCTTAT      663
QY      84      SerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrAspLysAla-----Ala      101
Db      664      GGGCTAGAGGATGAC-----CAGCGTAGTATGGGCTATGACCTGGATGATATGGT      714
QY      102      SerLysThrHisLeuSerAlaSerAlaSerIleProAspThrAspArgMetTrpTyrSerIleGly      121
Db      715      ATGATGTCGTGATTATGGCACTGCCCGTTCGCACCTGGGACACCCCTCGACCTCGTC-GGC      773
QY      122      AlaThrTyrLysPheThrProAsnLeuSerValGlyPheAlaHisLeuArgGly      141
Db      774      GCCTCA-----GGTAAGGCAGCA-ATAAGGGGA      799
QY      142      LysLysLysHisPheValGluThrGlnAsnIleLysGly      154
Db      800      AGAAGAAAGGTCCTTCCCAAGACCAAGCAACAGGGGG      838

RESULT 9
B1547625      672 bp      mRNA      linear      EST 05-SEP-2001
LOCUS      603191761F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5263190 5',
DEFINITION      mRNA sequence.
ACCESSION      B1547625
VERSION      B1547625.1 GI:15434937
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgapbs-femail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11663 row: i column: 15
High quality sequence stop: 672.
FEATURES      source
Location/Qualifiers
1..672
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5263190"
/tissue_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site1: BamHI; Site2: SalI-XhoI (gtcggag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.5 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      146 a      172 c      209 g      145 t
ORIGIN
Alignment Scores:
Pred. No.:      3.3      Length:      672
Score:      84.00      Matches:      35
Percent Similarity:      40.16%      Conservative:      14
Best Local Similarity:      28.69%      Mismatches:      53
Query Match:      8.78%      Indels:      20
DB:      13      Gaps:      6

```









```

ORIGIN
Alignment Scores:
Pred. No.: 4.88 Length: 649
Score: 82.50 Matches: 40
Percent Similarity: 37.79% Conservative: 25
Best Local Similarity: 23.26% Mismatches: 60
Query Match: 8.62% Indels: 47
DB: 10 Gaps: 9

US-09-995-493-52 (1-179) x AW553447 (1-649)
Qy 5 ValLeuGlyProTyrIleGlyLysGly---SerLeuThrLeuLysLeuPro---AlaTyr 22
Db 590 CTGTTGCTTCCTATGTCAGCAGTGAGCTGAACATGATCATGCTCCAGATGAGCAC 531

Qy 23 TrpGluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAlaIleHisTyrSerTyrLys 42
Db 530 GTTGAACGTAGT-----ACAGTGGAAAAGGAAGTAACCTTAGCAG---AAA 489

Qy 43 TyrThrGluTrpSerArgPheLysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyr 62
Db 488 TTTATAGAGTGCACAGCGCTGCACAAGATG-----GACGGAGAGAGGTA 444

Qy 63 GluAlaPheThrLysLysGluGluTyrLysAspAsnSerArgPheAlaIleGlyThrThr 82
Db 443 GAAGTATTCTCCCAAAAGTTTAAGCTGGAGGAAT----- 408

Qy 83 TyrSerLeuAsnAspAlaLeu----- 89
Db 407 TATAACATGAACGATGCCCTTACAAGTTGGCGCATGACTGATGCTTTGGCGGAGGGCA 348

Qy 90 -----ThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThr 104
Db 347 GACTTTTCTGCAATGCTCTCCCAAGCAAGGCTTTTCTGTCTAAGTTGTGCATAAGGCC 288

Qy 105 HisLeuSerAlaSerIleProAspThrAspArgMetTrpTyrSerIleGly----- 121
Db 287 TTTGTGGAGGTTAATGAGGAGGCGCAGAGCGCTGCAGCTGTACAGCTGGCGATGACG 228

Qy 122 ---AlaThrLysLysPheThrProAsnLeuSerValAspValGlyPheAlaHisLeuArg 140
Db 227 GTGAGGTGCATGAGATTCACCTCCCGCTTCTGTCGCCACCCCTTCCTTTCTTC--- 171

Qy 141 GlyLysLysLysHisPheValGluThrGlnAsnIle 152
Db 170 -----ATTCCACCATGTTAAGACCAATGGAATT 144

RESULT 15
BJ041818 683 bp mRNA linear EST 07-DEC-2001
LOCUS
DEFINITION
laevis cDNA NIBB Mochii normalized Xenopus neurula library Xenopus
BJ041818
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
African clawed frog.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
REFERENCE
1 (bases 1 to 683)
AUTHORS
Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-I,T. and Kohara
,Y.
TITLE
Expressed genes in X. laevis embryo
JOURNAL
Unpublished (2001)
COMMENT
Contact: Tadasi Shin-I
Center for Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
source
Location/Qualifiers
1..683
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="XL027106"
/clone_lib="NIBB Mochii normalized Xenopus neurula
library"
/tissue_type="whole embryo"
/dev_stage="stage 15"
/note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Staging
according to Nieuwkoop and Faber. Library is subcloned
and was constructed by N. Garrett and A.M. Zorn,
(Wellcome/CRC Institute)."
BASE COUNT 176 a 171 c 167 g 167 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 5.29 Length: 683
Score: 82.50 Matches: 35
Percent Similarity: 40.58% Conservative: 21
Best Local Similarity: 25.36% Mismatches: 57
Query Match: 8.62% Indels: 25
DB: 13 Gaps: 6

US-09-995-493-52 (1-179) x BJ041818 (1-683)
Qy 4 GlyValLeuGlyProTyrIleGlyLysGlySerLeuThrLeuLysLeuProAlaTyrTrp 23
Db 242 GGTATCGTTGGTGTAT---GGTGCAGCTGACCGTACCAACCTGCAGAGAGCT--- 292

Qy 24 GluLeuSerGlyPheHisGlnLeuThrAspGlnTrpAla----- 36
Db 293 CAACCTCTTGGCAACGGTAAAAAAGCTGAACAGTGGCTACTGCTCTGAAGTACGACGCG 352

Qy 37 -----IleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGluLeuArgGly 54
Db 353 AACAACTCTACCTGGCGGCACTACGGTGAAACCCGTAACGCTACGCCGATCACAAT 412

Qy 55 LysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGluTyrLysAspAsn 74
Db 413 AAATTACAAACACCGCGGCTCCGCCAACAAACGCAAGAGCTTCTGTAGTTGCCGAA 472

Qy 75 SerArgPheAlaIleGlyThrTyrSerLeuAsnAspAlaLeuThrLeuArgAlaGly 94
Db 473 TACCAGTTCGATTTTCGGT-----CTGCGTCCGTC 502

Qy 95 LeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSerIleProAspThrAsp 114
Db 503 ATCCCTTACCAACAACTTAAGCCGAAAGAC-----GTAGAGGTATCGGTGATGTTGAT 556

Qy 115 ArgMet---TrpTyrSerIleGlyAlaThrTyrLysPheThrProAsnLeuSer 131
Db 557 CTGCTGAACCTACTTTGAAGTGGCGCAACCTACTACTTCAACAAACATGCTCC 610

RESULT 16
BM930833 744 bp mRNA linear EST 13-MAR-2002
LOCUS
DEFINITION
UI-E-EJ1-ajb-k-10-0-UI.r1 UI-E-EJ1 Homo sapiens cDNA clone
UI-E-EJ1-ajb-k-10-0-UI 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
EST.
Homo sapiens.
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 744)
AUTHORS
Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
97044477

```

## COMMENT

Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
Tissue Procurement: Dr. Gregg Hageman  
cDNA Library preparation by: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
Seq primer: M13 REVERSE.

## FEATURES

source

1. .744  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="UI-E-EJ1-aj0-k-10-0-UI"  
/tissue\_type="UI-E-EJ1"  
/tissue\_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid"  
/dev\_stage="fetal and adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/note="Organ: eye; Vector: p773-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; UI-E-EJ1 is a subcloned cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (d)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAAGA; lens, CGATTAGCG; eye anterior segment, AATGCCCATG; optic nerve, CCATTAGTC; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."  
220 a 99 c 153 g 271 t 1 others

BASE COUNT

ORIGIN

Alignment Scores:  
Pred. No.: 6.03 Length: 744  
Score: 82.50 Matches: 25  
Percent Similarity: 43.33% Conservative: 14  
Best Local Similarity: 27.78% Mismatches: 30  
Query Match: 8.62% Indels: 21  
DB: 14 Gaps: 3

US-09-995-493-52 (1-179) x BM30833 (1-744)

QY 84 SerLeuAsnAspAlaLeuThrLeuArgAlaGlyLeuAlaTyrrAspLysAlaLaSerLys 103  
Db 739 TCATTAATAAATAATACAGTGACTTTGGGTAGGA-----AAAGAAGTGCATTA 686  
QY 104 ThrHisLeuSerAlaSerIleProAspThrAspArgMetrrprrSerIleGlyAlaThr 123  
Db 685 CTTCAATTATGTTTCCCTCCACAAAGAGCGGCCACAATTT---GTACATCCCA 629  
QY 124 TyrLysPheThrProAsnLeuSerValAspValGlyPheAlaHisLeuArgGlyLysLys 143  
Db 628 TACAATACNCAATTCAATTATCTTTCAAAATACTATATACACAAAGGGGAAGAAA 569  
QY 144 LysHisPheValGluThrGlnAsnIleLysGlyLeuLeuValGluAlaAspTyrThr 163  
Db 568 -----TATACA 563  
QY 164 ThrLysAlaThrAlaAsnLeuTyrGlyLeu 173  
||| |||:||||: |||

Db 562 GCATGCTCTTTTGGTAACATATGAATTG 533

RESULT 17  
AZ573577

LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AZ573577  
319PVA06 Pv MBN #30 Plasmodium vivax genomic 3', DNA sequence.  
AZ573577  
GI:13987813  
GSS.  
Plasmodium vivax.  
malaria parasite P. vivax.  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
1 (bases 1 to 665)  
Carlton,J.M.-R. and Dame,J.B.  
The Plasmodium vivax and P. berghei gene sequence tag projects  
Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)  
Contact: Dame JB  
Dept. of Pathobiology, College of Veterinary Medicine  
University of Florida  
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA  
Tel: 352 392 4700  
Fax: 352 392 9704  
Email: damej@mail.vetmed.ufl.edu  
Seq primer: M13(-20) forward  
Class: shotgun.

## FEATURES

source

Location/Qualifiers

1. .665  
/organism="Plasmodium vivax"  
/strain="Salvador I (Collins, W. 1972. J. Parasitol. 69, 497-598)"  
/db\_xref="taxon:5855"  
/clone\_lib="Pv MBN #30"  
/dev\_stage="asexual blood forms"  
/lab\_host="Salmiri boliviensis"  
/note="Vector: pBluescript SK(+) vector DNA, phagemid excised from lambda ZAP; Site\_1: EcoR V; Site\_2: EcoR V; Host leukocytes were extracted from P. vivax infected blood using the following methods: first, infected blood was activated by the addition of 0.5 ml of ADP (40mg/ml) per 10 ml blood. Then blood was passed over a column of acid washed 0.1 mm glass beads, then through a Plasmidipur filter, followed by passage through a column of pre-wet Whatman CF11 powder (1:2 ratio volume of blood to CF11), and finally centrifuged through a 50% Percoll density cushion. Purified DNA was digested with mung bean nuclease in the presence of 4% formaldehyde at 50°C as described (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). Digested DNA was blunt-ended using T4 DNA polymerase and size fractionated over a Sepharose CL-2B column. Fractions in the size range 500bp-4kb were ligated into the Eco RV site of pBluescript SK(+), and E. coli XL-10 Gold transformed with the ligation mixture."  
205 a 168 c 158 g 134 t

BASE COUNT

ORIGIN

Alignment Scores:  
Pred. No.: 5.88 Length: 665  
Score: 82.00 Matches: 35  
Percent Similarity: 39.46% Conservative: 23  
Best Local Similarity: 23.81% Mismatches: 49  
Query Match: 8.57% Indels: 40  
DB: 17 Gaps: 6

US-09-995-493-52 (1-179) x AZ573577 (1-665)

QY 50 LysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGlu 69  
||||| :|||:|||||  
Db 39 AAAGAAGCGTCAGATGTGAACCGAGATGAAGCAGCAGCTCCCGCTTCCCAATATGAAG 98  
QY 70 -----GlutylLysAspAsnSerArgPheAla---IleGly 80  
||| :|||:||||: |||  
Db 99 TACGGCAATATTTCTTATGCTATGAGGTGAGGAACAGAAATAGTTTTCAGACCTCTAC 158





```

DEFINITION
sej1a04.y1 Gm-cl072 Glycine max cDNA clone SOYBEAN CLONE ID:
Gm-cl072-5000 5' similar to TR:Q9SW37 Q9SW37 HYPOTHETICAL 61.4 KD
PROTEIN. ;, mRNA sequence.
BM178271
BM178271.1 GI:17401489
EST.
SOURCE
soybean.
ORGANISM
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE
1 (bases 1 to 563)
Shoenaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
TITLE
Public Soybean EST Project
JOURNAL
Unpublished (1999)
COMMENT
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@waterston.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact: ccu@resgen.com web site:
www.resgen.com
Seq primer: -40RP from Gibco
High quality sequence stop: 422.
FEATURES
Location/Qualifiers
1..563
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl072-5000"
/clone_lib="Gm-cl072"
/tissue_type="seedlings induced for symptoms of SDS
(Sudden Death Syndrome) disease"
/dev_stage="2-3 weeks old"
/lab_host="DH10B"
/note="Vector: pBluescript II SK+; Site.1: ECORI; Site.2:
XhoI; The cDNA library was constructed from mRNA isolated
from 2-3 week old seedlings that were induced for symptoms
of SDS (Sudden Death Syndrome) disease by the
translocation of Culture Filtrate of Fusarium solani f.
sp. glycinis (Plant Cell Report 18:375-380). Cultivar p.
567374 is partially resistant to the disease SDS. Plant
tissue (expanded leaves, folded leaves, and new shoots)
were collected at 1, 6, 24, and 48 hrs. after inoculation
and their mRNA pooled equally for cDNA construction. The
library was prepared using the Stratagene pBluescript II
SK(+) library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with an XhoI restriction site. EORI
adaptors were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA insert is protected
from XhoI digestion via methylation during first strand
synthesis. The cDNA fragments were directionally cloned
into the EORI-XhoI restriction site of the pBluescript
vector. The ligated cDNA fragments were transformed into
E.coli ElectroMax DH10B host cells. Plants were inoculated
by Shuxian Li (Glen Hartman lab, University of Illinois).
Library was constructed by Steve Clough (Gila Vodkin lab,
University of Illinois)."
174 a 99 c 102 q 188 t
BASE COUNT

```



```

source      1. .659
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODM002YF06"
/clone_lib="LTI_NFL001_NBC4"
/lab_host="NCI_CGAP_Skn3"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 165 a 220 c 199 g 75 t
ORIGIN

Alignment Scores:
Pred. No.: 7.81 Length: 659
Score: 81.00 Matches: 41
Percent Similarity: 41.01% Conservatives: 16
Best Local Similarity: 29.50% Mismatches: 44
Query Match: 8.46% Indels: 38
DB: 12 Gaps: 9

US-09-995-493-52 (1-179) x BG740114 (1-659)

Qy 38 HisTyrSerTyrLys-----TyrThrGluTrpSerArgPhe 49
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 97 CACTACACGCTCAAGGCTCAGTTCAAGACGTCCTCACTCTACACGGCGGGCGGCTGTTC 156
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 50 LysGluLeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGlu 69
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 157 AAGGAG-----AGCTGGCGGGCGGGCGGCGGAGGCGACACGGCGCGC--- 204
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 70 GluTyrLysAspAsnSerArgPheAlaIleGlyThrTyrThrTyrSerLeuAsnAlaLeu 89
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 205 -----CTGACGGCGCTCGCATCGCAGATCTCGGAGCTCGCTCGCTCGGC----- 249
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 90 ThrLeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSerAlaSer 109
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 250 -----CCGAGGGG---GCCACGACAAAGGCGGGCGGCGCT-----GCCGCC 288
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 110 IleProAspThrAspArgMetTrpTyrSerIleGly---AlaThrTyrLys----- 125
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 289 GTCTCAGACGAGCTCGGCGGCTGATACAGGCTTCCACCGGCTCGCACAAGGACACAGC 348
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 126 -----PheThrProAsnLeuSerValAspValGlyPhe----- 136
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 349 CGCCTGTGCGACACAGCTCCACCTCTCTCGGACAGCGGCTCGCAAGTACAGCACCTCCTC 408
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 137 ---AlaHisLeuArgGlyLysLysLysHisPheValGluThrGlnAsnIleLysGly 154
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 409 CCAGAGCACCTTCGTGGCGCACAGCAGGGTCACCAGGATGCCCGACAGATGTGCAAGGG 465
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

RESULT 24
LOCUS      AL531424 873 bp mRNA linear EST 13-FEB-2001
DEFINITION AL531424 LTI_NFL001_NBC4 Homo sapiens cDNA clone CSODM002YF06 5
prime, mRNA sequence.
ACCESSION  AL531424
VERSION    AL531424.1 GI:12794917
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 873)
AUTHORS   Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES   Location/Qualifiers
source     1. .873

```

---

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODM002YF06"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT 187 a 221 c 254 g 209 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 12.1 Length: 873
Score: 81.00 Matches: 34
Percent Similarity: 40.16% Conservatives: 15
Best Local Similarity: 27.87% Mismatches: 53
Query Match: 8.46% Indels: 20
DB: 9 Gaps: 6

US-09-995-493-52 (1-179) x AL531424 (1-873)

Qy 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 98 ACTGTGAGATGGTGTGTCAGTGTGCGGAGCATGAGGCACATAAGTGCACAGATTTCGCGAC 157
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 52 ---LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGlu 70
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 158 CATATGAAGAAGCGTCATTCCATCCGATGTCCTGCTGTGTGAAGAAATCCTCC 217
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 71 TyrLysAspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThr 90
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 218 TACCTTGATTCATCAGCAG---GCCATTGCGCAAGGAGCG-----GATGCTGTGACA 268
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 91 LeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSer----- 107
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 269 CTGGATGCAAGTTGGTGTATGATGCTTACCTGGCTCCCAATACCTGAAGCTGTGGTG 328
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 108 -----AlaSerIleProAspThrAspArgMetTrpTyrSerIleGlyAlaThr 123
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 329 GCAGAGTCTATGGGTCAAAAGAGGATCCACAGACTTCTATTATGCTGTGCTGTGGTG 388
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 124 TyrLysPheThrProAsnLeuSerValAspValGlyPheAla-----HisLeuArgGly 141
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 389 AAGAAG-----GATAGTGGCTTCCAGATGACACGAGTTCGAGGC 427
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

RESULT 25
LOCUS      W64749 484 bp mRNA linear EST 10-JUN-1996
DEFINITION W64749 Soares mouse embryo-NDMEJ3.5 14.5 Mus musculus cDNA
clone IMAGE:385062 5' similar to PIR:A47411 A47411
ADPrribosylarginine hydrolase ;, mRNA sequence.
ACCESSION  W64749
VERSION    W64749.1 GI:1372452
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1 (bases 1 to 484)
AUTHORS   Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

```











```

Db 261 CTGGATGCAGGTTTGGTGTATGATGCTTACCTGGCTCCCAATAAACCTGAAGCCTCTGGTG 320
Qy 108 -----AlaSerIleProAspThrAspArgMetTrpTyrSerIleGlyAlaThr 123
Db 321 GCAGAGTCTTATGGGTCAAAAGAGATCCACAGACTTCTATTATGCTGTGCTGGTG 380
Qy 124 TyrLysPheThrProAsnLeuSerValAspValGlyPheAla-----HisLeuArgGly 141
Db 381 AAGAAG-----GATAGTGCTTCCAGATCAACAGCTTCGAGGC 419
Qy 142 LysLys 143
Db 420 AAGAAG 425

RESULT 32
LOCUS BG700910 602681855F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4814772 5',
DEFINITION mRNA sequence.
ACCESSION BG700910
VERSION BG700910.1 GI:13970725
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 615)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10710 row: m column: 13
High quality sequence stop: 612.
FEATURES
Location/Qualifiers
1..615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4814772"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
size-selected for average insert size 2.5 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 129 a 156 c 185 g 145 t
ORIGIN

Alignment Scores:
Pred. No.: 9.44 Length: 615
Score: 80.00 Matches: 34
Percent Similarity: 40.16% Conservative: 15
Best Local Similarity: 27.87% Mismatches: 53
Query Match: 8.36% Indels: 20
DB: 12 Gaps: 6

US-09-995-493-52 (1-179) x BG700910 (1-615)
Qy 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51

```

```

Db 121 ACTGTGAGATGCTGTCAGTGTGGAGCATGAGGCCACTAAGTGCAGAGTTTCCGCGAC 180
Qy 52 ---LeuArgGlyLysTyrClnAspGlySerGlyTyrGluAlaPheThrLysLysGluGlu 70
Db 181 CATATGAAGAAGCGTCATTCCATCCGATGGTCCCGATGGTGTGTGTAAGAAGACCTCC 240
Qy 71 TyrLysAspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThr 90
Db 241 TACCTTGATTGATCAGG---GCCATTGGCGCAACGAGCG-----GATGCTGTGACA 291
Qy 91 LeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSer----- 107
Db 292 CTGGATGCAGGTTTGGTGTATGATGCTTACCTGGCTCCCAATAAACCTGAAGCCTGGTG 351
Qy 108 -----AlaSerIleProAspThrAspArgMetTrpTyrSerIleGlyAlaThr 123
Db 352 GCAGAGTCTTATGGGTCAAAAGAGATCCACAGACTTCTATTATGCTGTGCTGGTG 411
Qy 124 TyrLysPheThrProAsnLeuSerValAspValGlyPheAla-----HisLeuArgGly 141
Db 412 AAGAAG-----GATAGTGCTTCCAGATCAACAGCTTCGAGGC 450
Qy 142 LysLys 143
Db 451 AAGAAG 456

RESULT 33
LOCUS BQ639058 617 bp mRNA linear EST 15-JUL-2002
DEFINITION hd30h04.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he
ACCESSION BQ639058
VERSION BQ639058.1 GI:21763517
KEYWORDS EST.
SOURCE Homo sapiens human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 617)
AUTHORS Wistow,G., Bernstein,S.L., Wyatt,M.K., Ray,S., Behal,A., Touchman
,J.W., Bouffard,G., Smith,D. and Peterson,K.
TITLE Expressed sequence tag analysis of human retina for the NEIBank
Project: Retbindin, an abundant, novel retinal cDNA and alternative
splicing of other retina-preferred gene transcripts
JOURNAL Mol. Vis. 8 (4). (2002) In press
COMMENT Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 30 row: h column: 04
Seq primer: M13RP1 reverse primer (ABI).
FEATURES
Location/Qualifiers
1..617
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hd30h04"
/clone_lib="Human Retina cDNA (Un-normalized, unamplified
): hd/he"
/tissue_type="Retina"
/dev_stage="Adult"
/lab_host="EMDH10B"
/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue
was dissected from two 80 year old donors with no observed
eye disease. 100ug of total RNA was used for library
construction. A directionally cloned cDNA library in the
pSPORT1 vector (Life Technologies) was constructed at
Bioserve Biotechnology (Laurel MD) essentially following
the protocols of the SuperScript Plasmid System full
details of which are contained in the manufacturer's

```











Qy 108 -----AlaSerIleProAspThrAspArgMetTrpTyrSerIleGlyAlaThr 123  
 Db 333 GCAGAGTTCTATGGTCAAAAGAGATCCACAGACTTCTATTATGCTGTTGCTGGTG 392  
 Qy 124 TyrLysPheThrProAsnLeuSerValAspValGlyPheAla-----HisLeuArgGly 141  
 Db 393 AAGAAG-----GATAGTGGCTTCCAGATGAACACGCTTCGAGGC 431  
 Qy 142 LysLys 143  
 Db 432 AAGAAG 437  
 RESULT 40  
 BE006001/c  
 LOCUS BE006001 665 bp mRNA linear EST 05-JUN-2000  
 DEFINITION RC0-BN0121-280300-032-b06 BN0121 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BE006001  
 VERSION BE006001.1 GI:8266234  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 665)  
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 M.J., Soares,F., Brentani,R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.  
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RC0-BN0121-280  
 300-032-b06&t3-2000-03-28&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 23  
 High quality sequence stop: 664.  
 FEATURES  
 source Location/Qualifiers  
 1..665  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="BN0121"  
 /dev\_stage="Adult"  
 /note="Organ: breast\_normal; Vector: puc18; Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 BASE COUNT 169 a 188 c 159 g 149 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 10.7 Length: 665  
 Score: 80.00 Matches: 34  
 Percent Similarity: 40.16% Conservative: 15  
 Best Local Similarity: 27.87% Mismatches: 53  
 Query Match: 8.36% Indels: 20  
 DB: 10 Gaps: 6

US-09-995-493-52 (1-179) x BE006001 (1-665)  
 Qy 32 ThrAspGlnTrpAlaIleHisTyrSerTyrLysTyrThrGluTrpSerArgPheLysGlu 51  
 Db 646 ACTGTGAGATGGTGTGCGAGATGCGAGCCCACTAAGTCCACAGAGCTTCCGCGAC 587  
 Qy 52 ---LeuArgGlyLysTyrGlnAspGlySerGlyTyrGluAlaPheThrLysLysGluGlu 70  
 Db 586 CATATGAAGCGTCATTCATCCATCCGATGGTCCCGAGTCTTCTGTGTGAAGAACCTCC 527  
 Qy 71 TyrLysAspAsnSerArgPheAlaIleGlyThrThrTyrSerLeuAsnAspAlaLeuThr 90  
 Db 526 TACCTTGATTGCATCAGG---GCCATTGCGGCAACGAAGCG-----GATGCTGTGACA 476  
 Qy 91 LeuArgAlaGlyLeuAlaTyrAspLysAlaAlaSerLysThrHisLeuSer----- 107  
 Db 475 CTGGATCGAGTTTGGTGTATGATGCTTACTGTGCTCCCAATAACCTGAAGCCTGTGGTG 416  
 Qy 108 -----AlaSerIleProAspThrAspArgMetTrpTyrSerIleGlyAlaThr 123  
 Db 415 GCAGAGTTCTATGGTCAAAAGAGATCCACAGACTTCTATTATGCTGTGCTGGTG 356  
 Qy 124 TyrLysPheThrProAsnLeuSerValAspValGlyPheAla-----HisLeuArgGly 141  
 Db 355 AAGAAG-----GATAGTGGCTTCCAGATGAACACGCTTCGAGGC 317  
 Qy 142 LysLys 143  
 Db 316 AAGAAG 311

Search completed: May 19, 2003, 19:57:24  
 Job time : 1088 secs

GenCore version 5.1.4\_p5\_4578.  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 09:39:39 ; Search time 15 Seconds  
(without alignments)  
351.114 Million cell updates/sec

Title: US-09-995-493-52

Perfect score: 957

Sequence: 1 QHNGVLGYIGKSLTLKP.....ADYTKATANLYGLNLNRYF 179

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	559	58.4	455	2	US-08-472-172-4
2	549	57.4	459	2	US-08-472-172-2
3	502.5	52.5	432	2	US-08-472-172-6
4	252	26.3	77	2	US-08-472-172-12
5	129	13.5	34	1	US-08-475-989-8
6	129	13.5	34	2	US-08-475-985-8
7	129	13.5	34	2	US-08-472-172-20
8	129	13.5	34	3	US-08-256-839-8
9	111	11.6	32	1	US-08-475-989-10
10	111	11.6	32	2	US-08-475-985-10
11	111	11.6	32	2	US-08-472-172-22
12	111	11.6	32	3	US-08-256-839-10
13	89	9.3	459	2	US-08-810-655A-2
14	87.5	9.1	35	1	US-08-475-989-14
15	87.5	9.1	35	2	US-08-475-985-14
16	87.5	9.1	35	2	US-08-472-172-26
17	87.5	9.1	35	3	US-08-256-839-14
18	85.5	8.9	340	5	PCT-US95-16126-1
19	85.5	8.9	340	5	PCT-US95-16126-1
20	83.5	8.7	261	4	US-08-961-083-152
21	83.5	8.7	362	1	US-08-183-214-2
22	82	8.6	54	1	US-08-475-989-49
23	82	8.6	54	2	US-08-475-985-49
24	82	8.6	54	3	US-08-256-839-49
25	82	8.6	188	1	US-08-486-715-5
26	82	8.6	188	1	US-08-486-719-5
27	82	8.6	188	1	US-08-476-100-5

28	82	8.6	188	3	US-08-475-749-5	Sequence 5, Appl
29	81.5	8.5	362	1	US-08-183-214-10	Sequence 10, Appl
30	81	8.5	188	1	US-08-286-767-3	Sequence 3, Appl
31	80	8.4	38	1	US-08-475-989-13	Sequence 13, Appl
32	80	8.4	38	2	US-08-475-985-13	Sequence 13, Appl
33	80	8.4	38	3	US-08-472-172-25	Sequence 25, Appl
34	80	8.4	38	3	US-08-256-839-13	Sequence 13, Appl
35	80	8.4	53	1	US-08-475-989-48	Sequence 48, Appl
36	80	8.4	53	2	US-08-475-985-48	Sequence 48, Appl
37	80	8.4	53	3	US-08-256-839-48	Sequence 48, Appl
38	80	8.4	55	1	US-08-475-989-47	Sequence 47, Appl
39	80	8.4	55	2	US-08-475-985-47	Sequence 47, Appl
40	80	8.4	55	3	US-08-256-839-47	Sequence 47, Appl
41	80	8.4	164	4	US-09-370-838-203	Patent No. 5262177
42	80	8.4	696	6	5262177-4	Sequence 2, Appl
43	80	8.4	698	2	US-08-175-158A-2	Sequence 2, Appl
44	80	8.4	1074	2	US-08-470-058-2	Sequence 2, Appl
45	80	8.4	1074	3	US-09-037-188-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1  
US-08-472-172-4  
; Sequence 4, Application US/08472172  
; Patent No. 5985288  
; GENERAL INFORMATION:  
; APPLICANT: Munson, Jr., Robert S  
; APPLICANT: Grass, Susan  
; APPLICANT: Chong, Pele Y Y  
; APPLICANT: Fahim, Raafat  
; APPLICANT: Sia, Charles D Y  
; APPLICANT: McVerry, Patrick  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Outer Membrane Protein p1 and Peptides  
; TITLE OF INVENTION: of Haemophilus Influenzae Type B  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R6  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/472,172  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/849,411  
; FILING DATE: 07-JUL-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-471  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; TELEX: 063-24567 SIMBAS  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 455 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-08-472-172-4



Db 341 SNNRIALGASYNLDEKLTLAGIAYDOASRHRSAIIPDTRWYSLGATYKFTPNLS 400  
QY 132 VDVGFAHLRGKKHVFETONIK-GLLLVBEADY 162  
Db 401 VDLGYAYLKGKKVHFKEAQAAAGGFTTTANY 432

## RESULT 4

US-08-472-172-12  
; Sequence 12, Application US/08472172  
; Patent No. 5985288  
; GENERAL INFORMATION:  
; APPLICANT: Munson, Jr., Robert S  
; APPLICANT: Grass, Susan  
; APPLICANT: Chong, Pele Y Y  
; APPLICANT: Fahim, Raafat  
; APPLICANT: Sia, Charles D Y  
; APPLICANT: McVertry, Patrick  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Outer Membrane Protein P1 and Peptides  
; TITLE OF INVENTION: Of Haemophilus Influenzae Type B  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R6  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/472,172  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/849,411  
; FILING DATE: 07-JUL-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-471  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; TELEX: 063-24567 SIMBAS  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 77 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-472-172-12

Query Match 26.3%; Score 252; DB 2; Length 77;

Best Local Similarity 64.9%; Pred. No. 2.9e-20;

Matches 50; Conservative 9; Mismatches 14; Indels 4; Gaps 1;

QY 107 SASIPDTRMWYSIGATYKFTPNLSVDVGFAHLRGKKHVFETONIKG---LLLVBEADY 162

Db 1 SAAIPDTRTWYSLGATYKFTPNLSVDLGAYLKGKKVHFKEVKTIGDKRTRLNTTANY 60

QY 163 TTKATANLYGLNLNRYF 179

Db 61 TSOAHANLYGLNLNYSF 77

## RESULT 5

US-08-475-989-8

; Sequence 8, Application US/08475989  
; Patent No. 5679352  
; GENERAL INFORMATION:  
; APPLICANT: CHONG, Pele  
; APPLICANT: KANDIL, Ali  
; APPLICANT: Sia, Charles  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Synthetic Haemophilus Influenzae  
; TITLE OF INVENTION: Conjugate Vaccine  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/475,989  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/256,839  
; FILING DATE: 03-FEB-1993  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA93/00041  
; FILING DATE: 03-FEB-1993  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9202219.3  
; FILING DATE: 03-FEB-1992  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, MICHAEL I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-505 MIS:vg  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 34 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-475-989-8

Query Match 13.5%; Score 129; DB 1; Length 34;

Best Local Similarity 66.7%; Pred. No. 2.1e-07;

Matches 22; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 24 ELSGFHQLTDQWAIHYSYKYTWSRKFELRGKY 56

Db 2 ELSGFHQLTDKILAVHYSYKYTHWSRLTKLHASF 34

## RESULT 6

US-08-475-985-8

; Sequence 8, Application US/08475985

; Patent No. 5972349

; GENERAL INFORMATION:

; APPLICANT: CHONG, Pele

; APPLICANT: KANDIL, Ali

; APPLICANT: Sia, Charles

; APPLICANT: Klein, Michel

; TITLE OF INVENTION: Synthetic Haemophilus Influenzae



RESULT: TRUE

Db 1 LYEKLTRAGIAYDQASRHHRSAAIPDTR 31

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```



```

; TELEPHONE: (716) 856-4000
; TELEFAX: (716) 849-0349
; INFORMATION FOR SEQ ID NO: 2
; SEQUENCE CHARACTERISTICS:
; LENGTH: 459 residues
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Moraxella catarrhalis
; FEATURE:
; OTHER INFORMATION: leader
US-08-810-655A-2

```

```

Query Match      9.3%; Score 89; DB 2; Length 459;
Best Local Similarity 23.2%; Pred. No. 0.18;
Matches 42; Conservative 19; Mismatches 84; Indels 36; Gaps 6;

QY 9 YIGKSLTLKLPAYWELSGFHQLDQWAIHYSYKTEWSREKELRGKYODSGYEAFKK 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 292 YAKKDKFVTLPDSNNLDFQGVNPTLLTAKVRVPMSEDFIRPTQYTE-----TTK 344

QY 69 EY-----KDNSEFAIGTYSYLNDAITLRAGLAYDKAAKTHLSASIPDTRMWY 118
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 345 LRYPQGLPIISYDXDQWSAEYGLGKRYSRDLAVSGAVGWDSCAGN---PASSLGPICKY 401

QY 119 S--IGATYKFTPNLSVDVGFALHKGKKHFFVETQNIKGLLLVEADYTTKATANLYGLNL 176
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 402 SLGLGARYNYTPWSLSLG-----GKYFKGDAQ-----AQLPTKDKVGNPDSNDG 447

QY 177 Y 177
   |||
Db 448 Y 448

```

```

RESULT 14
US-08-475-989-14
: Sequence 14, Application US/08475989
: Patent No. 5679352
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: KANDIL, Ali
: APPLICANT: SIA, Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: Synthetic Haemophilus Influenzae
: TITLE OF INVENTION: Conjugate Vaccine
: NUMBER OF SEQUENCES: 56
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Slim & McBurney
: STREET: Suite 701, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/475,989
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/256,839
: FILING DATE: 03-FEB-1993
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/CA93/00041
: FILING DATE: 03-FEB-1993
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9202219.3

```

```

; FILING DATE: 03-FEB-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-505 MIS:vq
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-475-989-14

Query Match          9.18; Score 87.5; DB 1; Length 35;
Best Local Similarity 54.3%; Pred. No. 0.0069;
Matches 19; Conservative 4; Mismatches 11; Indels

QY 146 FVETONIK-GLLLV EADYTTRATNLYGLNLNYRF 179
      ||| | :||| :|||||
DB 1 FKEAQQAGGFTTANTYSOAHANLYGLNLNYP 35
      ||| | :||| :|||||

RESULT 35
US-08-475-985-14
; Sequence 14, Application US/08475985
; Patent No. 5972349
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: KANDIL, Ali
; APPLICANT: SIA, Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: Synthetic Haemophilus Influenzae
; TITLE OF INVENTION: Conjugate Vaccine
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,985
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA93/00041
; FILING DATE: 03-FEB-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9202219.3
; FILING DATE: 03-FEB-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-506 MIS:vq
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163

```

## ; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 35 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-475-985-14

Query Match 9.1%; Score 87.5; DB 2; Length 35;  
Best Local Similarity 54.3%; Pred. No. 0.0069;  
Matches 19; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

## Qy 146 FVETQNIK-GLLLVEADYTTKATNLYGLNLYRF 179

Db 1 FKEAQAAGGFITTTANYTSQAHANLYGLNLYSF 35

## RESULT 16

US-08-472-172-26

; Sequence 26, Application US/08472172

; Patent No. 5985288

; GENERAL INFORMATION:

; APPLICANT: Munson, Jr., Robert S

; APPLICANT: Grass, Susan

; APPLICANT: Chong, Pele Y Y

; APPLICANT: Fahim, Raafat

; APPLICANT: Sia, Charles D Y

; APPLICANT: McVerry, Patrick

; APPLICANT: Klein, Michel

; TITLE OF INVENTION: Outer Membrane Protein PI and Peptides

; TITLE OF INVENTION: Of Haemophilus Influenzae Type B

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: Suite 701, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R6

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/472.172

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/849,411

; FILING DATE: 07-JUL-1992

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Stewart, Michael I

; REGISTRATION NUMBER: 24973

; REFERENCE/DOCKET NUMBER: 1038-471

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1163

; TELEX: 063-24567 SIMBAS

; INFORMATION FOR SEQ ID NO: 26:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 35 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-472-172-26

## Query Match

Best Local Similarity 9.1%; Score 87.5; DB 2; Length 35;

Matches 19; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

## Qy 146 FVETQNIK-GLLLVEADYTTKATNLYGLNLYRF 179

Db 1 FKEAQAAGGFITTTANYTSQAHANLYGLNLYSF 35

## RESULT 17

US-08-256-839-14

; Sequence 14, Application US/08256839

; Patent No. 6018019

; GENERAL INFORMATION:

; APPLICANT: CHONG, Pele

; APPLICANT: KANDIL, Ali

; APPLICANT: SIA, Charles

; APPLICANT: KLEIN, Michel

; TITLE OF INVENTION: Synthetic Haemophilus Influenzae

; TITLE OF INVENTION: Conjugate Vaccine

; NUMBER OF SEQUENCES: 56

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: Suite 701, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/256,839

; FILING DATE:

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: STEWART, MICHAEL I.

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-373 MIS:jb

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1163

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 35 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-256-839-14

Query Match 9.1%; Score 87.5; DB 3; Length 35;

Best Local Similarity 54.3%; Pred. No. 0.0069;

Matches 19; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

## Qy 146 FVETQNIK-GLLLVEADYTTKATNLYGLNLYRF 179

Db 1 FKEAQAAGGFITTTANYTSQAHANLYGLNLYSF 35

## RESULT 18

US-08-355-844-1

; Sequence 1, Application US/08355844

; Patent No. 5940307

; GENERAL INFORMATION:

; APPLICANT: Fischbarg, Jorge

; APPLICANT: Czegledy, Ferenc

; APPLICANT: Iserovich, Pavel

; APPLICANT: Li, Jun

; APPLICANT: Cheung, Min

; TITLE OF INVENTION: A METHOD FOR PREDICTING PROTEIN

; TITLE OF INVENTION: STRUCTURE

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond

; STREET: 30 Rockefeller Plaza

; CITY: New York

```
STATE: NY
COUNTRY: USA
ZIP: 10112-0228
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/355,844
FILING DATE: 14-DEC-1994
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Tang, Henry Y.S.
REGISTRATION NUMBER: 29,705
REFERENCE/DOCKET NUMBER: A39927-50/29910.
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-408-2586
TELEFAX: 212-765-2519
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 340 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..340
OTHER INFORMATION: OmpF porin protein
US-08-355-844-1

Query Match 8.9%; Score 85.5; DB 2; Length 340;
Best Local Similarity 25.2%; Pred. No. 0.28;
Matches 36; Conservative 22; Mismatches 60; Indels 25; Gaps 6;

QY 4 GVLGPYIGKSLTKLPAYWELSGFHLTDQW-----IHYSKYTWSRRELRC 54
Db 186 GIVGAY-GAADRTNLQEA--QPLNGKKAOWATGLKYDANNIYLAANYGETRNATPIIN 242
QY 55 KYDQSGCYEAFYKKEEYKNSRFAIGTYSLNDAITLRLAGLAYDKAASKTHLSASIPDPTD 114
Db 243 KFTNTSGFANKTQDVLVLAQYQDFG-----LRPSIATYKSKARD--VEGICDVD 290
QY 115 RM-WYSIGATYKFTPNLSVDVGF 136
Db 291 LVNYFVGATYTFNKNMSTYVDY 313

RESULT 19
PCT-US95-16126-1
Sequence 1, Application PC/TUS9516126
GENERAL INFORMATION:
APPLICANT: Fischbarg, Jorge
APPLICANT: Czegledy, Ferenc
APPLICANT: Iserovich, Pavel
APPLICANT: Li, Jun
APPLICANT: Cheung, Min
TITLE OF INVENTION: A METHOD FOR PREDICTING PROTEIN
TITLE OF INVENTION: STRUCTURE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10112-0228
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/355,844
FILING DATE: 14-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Tang, Henry Y.S.
REGISTRATION NUMBER: 29,705
REFERENCE/DOCKET NUMBER: A39927-50/29910
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-408-2586
TELEFAX: 212-765-2519
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 340 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..340
OTHER INFORMATION: OmpF porin protein
PCT-US95-16126-1

Query Match 8.9%; Score 85.5; DB 5; Length 340;
Best Local Similarity 25.2%; Pred. No. 0.28;
Matches 36; Conservative 22; Mismatches 60; Indels 25; Gaps 6;

QY 4 GVLGPYIGKSLTKLPAYWELSGFHLTDQW-----IHYSKYTWSRRELRC 54
Db 186 GIVGAY-GAADRTNLQEA--QPLNGKKAOWATGLKYDANNIYLAANYGETRNATPIIN 242
QY 55 KYDQSGCYEAFYKKEEYKNSRFAIGTYSLNDAITLRLAGLAYDKAASKTHLSASIPDPTD 114
Db 243 KFTNTSGFANKTQDVLVLAQYQDFG-----LRPSIATYKSKARD--VEGICDVD 290
QY 115 RM-WYSIGATYKFTPNLSVDVGF 136
Db 291 LVNYFVGATYTFNKNMSTYVDY 313

RESULT 20
US-08-961-083-152
Sequence 152, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
```

APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 152:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 261 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-961-083-152

Query Match 8.7%; Score 83.5; DB 4; Length 261;  
Best Local Similarity 23.9%; Pred. No. 0.32;  
Matches 38; Conservative 26; Mismatches 68; Indels 27; Gaps 7;  
QY 8 PYIGKSLTLKLPAYWELSGFHQLTQDQWATHYSYK---YTEWSRFKELCKYODGSGYEA 64  
DB 8 PFPKSGTIGPYNGIRLEGPK--SEW--YFDKNGVLQEFYGVGKTLKTDKDSVGRKY 63  
QY 65 FTKKEEYKDNRSFAIGTYSLNDAITLRAGLAYDKA----ASKTHLS-----ASI 110  
DB 64 GKRDESEDEKERYNYNYFNQHSLETGWLQDSNNWYLLAKTEINGENYLGERRAGW 123  
QY 111 PTDORMWYSGATKFTPNLSVDVGFHLRGGKKHVFET 149  
DB 124 INDDSTWYLDPTGTI-----MOTGWOYL-GNKWYLYRS 156

RESULT 21  
US-08-183-214-2  
Sequence 2, Application US/08183214  
Patent No. 5716816  
GENERAL INFORMATION:  
APPLICANT: Moss, Joel  
APPLICANT: Stanley, Sally J.  
APPLICANT: Nightingale, Maria S.  
APPLICANT: Murtagh, Jr., James J.  
APPLICANT: Monaco, Lucia  
APPLICANT: Takada, Tatsuyuki  
TITLE OF INVENTION: CLONES ENCODING MAMMALIAN  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend  
STREET: One Market Plaza, Steuart Street Tower  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/183,214  
FILING DATE: 14-JAN-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/888,231  
FILING DATE: 22-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Weber, Kenneth A.  
REGISTRATION NUMBER: 31,677  
REFERENCE/DOCKET NUMBER: 15280-27  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-543-9600  
TELEFAX: 415-543-5043  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 362 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-183-214-2

Query Match 8.7%; Score 83.5; DB 1; Length 362;  
Best Local Similarity 26.7%; Pred. No. 0.5;  
Matches 36; Conservative 19; Mismatches 55; Indels 25; Gaps 7;  
QY 11 GKGSLLTLKLP---AYWELSGFHQLTQDQWATHYSYKTEWSRFKELCKYODGSGYEAFTK 67  
DB 197 GKGLMEV-LPEAKAYVTQSGY--FVKENLQHWSEYFEWEKYLELRG-ILDGKSPVFPQ 252  
QY 68 KEYKDNRSFAIGTYS-----LNDAITLRAGLAYDKAASKTHLSASIPDPTD 114  
DB 253 PFGVKERDQFYEVYSYSGWSSGSHDAPMIAYDAL-LAAGDSNKLAAHRAFFHGGSDST 311  
QY 115 R-----MWSYSGATYK 125  
DB 312 ATTAGCWMVYHGFK 326

RESULT 22  
US-08-475-989-49  
Sequence 49, Application US/08475989  
Patent No. 5679352  
GENERAL INFORMATION:  
APPLICANT: CHONG, Pele  
APPLICANT: KANDIL, Ali  
APPLICANT: SIA, Charles  
APPLICANT: KLEIN, Michel  
TITLE OF INVENTION: Synthetic Haemophilus Influenzae  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,989  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/256,839  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA93/00041  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9202219.3  
FILING DATE: 03-FEB-1992  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, MICHAEL I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-505 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155

US-08-473-983-49

STREET: 225  
CITY: Boston

STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,715  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION NUMBER: 08/090,526  
FILING DATE: 09-JUL-1993  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION NUMBER: 07/529,602  
FILING DATE: 18-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00786/192005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 188 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-486-715-5

```

Query Match      8.5%; Score 82; DB 1; Length 188;
Best Local Similarity 23.7%; pred. No. 0.29;
Matches 45; Conservative 25; Mismatches 52; Indels 68; Gaps 11;

QY 32 TDQWAIHVS-YKWTWSEKRELRG-----KYQDGGSGVEATFK----- 67.
   | : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 25 TNAFSGCYARYAQSKVQDFNIRGVNKNRYEDDSVFTSSLSLYLGDROASGVPEEG 84
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 68 -----KEEYKDNSFEATGTTYSLNDATL---RAGLAYDKAASKTHLSASIPDTP----- 114
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 85 IHYHDKFEVXGSG-LMWGPAYRLSDNFSLYALAGVGTVKATFKEH---STQDGSFSNKI 140
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 115 ---RMWYSIGATYKFTP--NLSVDVDGFAHLRGKKKHFVEQTQIKGLLVEADYTTKATAN 169
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 141 SSRKTGFAGAGVQMNPENIVVDGY-----EGSNIS-----STK--IN 178
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 170 LYGLNLNRYF 179
   | : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 179 GFNVGVGYRF 188

```

RESULT 26  
US-0-486-719-5  
; Sequence 5, Application US/08486719  
; Patent No. 5674736  
; GENERAL INFORMATION:  
; APPLICANT: Miller, Samuel I.  
; TITLE OF INVENTION: SALMONELLA VIRULENCE GENES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P. C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

```

COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,719
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/090,526
FILING DATE: 09-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/629,602
FILING DATE: 18-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/192002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 188 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-486-719-5

```

```

Query Match      8.6%; Score 82; DB 1; Length 188;
Best Local Similarity 23.7%; Pred. No. 0.29;
Matches 45; Conservative 25; Mismatches 52; Indels 68; Gaps 11;

QY      32 TDQWATHYS-YKWTENSRFRELRG-----KYQDGSGEAFTK----- 67
      | : : : | : : | : : | : : | : : | : : | : : | :
DB      25 TNAFSGVYARYAOSKQVDFKNIIRGVNVKYRYEDDSPVSFISSLSLYGDRQASGSVEPEG 84
      | : : : | : : | : : | : : | : : | : : | : : | :
QY      68 -----KEEYKDNSRFAIGTTTYSLUNDALT--RAGLAYDKAAKTHLSASIPDTD----- 114
      | : : : | : : | : : | : : | : : | : : | : : | :
DB      85 IYHDKFEVKXGS-LWVGPAYRLSDNPSFLYALAGVGTVKATFREH---STQDGDGSFSNKI 140
      | : : : | : : | : : | : : | : : | : : | : : | :
QY      115 ---RMAYSIGATVKFPT--NLSVDVCGFAHLRGKKHVEVTQNLKGLLVEADYTTKATAN 169
      | : : : | : : | : : | : : | : : | : : | : : | :
DB      141 SSRKTGFAGGQVMQPLENIVVDVG!-----EGSNIS-----STK--IN 178
      | : : : | : : | : : | : : | : : | : : | : : | :
QY      170 YGLNLNRYR 179
      | : : | : : | : : | : : | : : | : : | : : | :
DB      179 GFNVGVGYRF 188
      | : : | : : | : : | : : | : : | : : | : : | :

```

```

RESULT 27
US-08-476-100-5
; Sequence 5, Application US/08476100
; Patent No. 5731196
; GENERAL INFORMATION:
; APPLICANT: Miller, Samuel I., III
; APPLICANT: Mekalanos, John J.
; TITLE OF INVENTION: SALMONELLA VIRULENCE GENES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P. C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; Zip: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:

```



;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/888,231  
;; FILING DATE: 22-MAY-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Weber, Kenneth A.  
;; REGISTRATION NUMBER: 31,677  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415-543-9600  
;; TELEFAX: 415-543-5043  
;; INFORMATION FOR SEQ ID NO: 10:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 362 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-183-214-10

Query Match 8.5%; Score 81.5; DB 1; Length 362;  
Best Local Similarity 27.2%; Pred. No. 0.83;  
Matches 37; Conservative 18; Mismatches 54; Indels 27; Gaps 8;  
QY 11 GKSLTLKLP---AYWELSGFHLTDQWAIHYSYKYTWSRFEKLRGKYQDGSYEATFK 67  
Db 197 GKGLMEV-LPEAKKVTOSGY--FVKENLOHWSYEKEKEKYLELRG-ILDGSAPVFPQ 252  
QY 68 KEKYKDNRSRFAIGTYS-----LNDALTLAGLAYDKAAASKTHLSASIPDTD 114  
Db 253 PFGVRRQFYIDVSYSGWSSGSHDPMIAYDAL-LAAGDSWKELAHAFPHGSDSDST 311  
QY 115 R---MWSYGATYKF 126  
Db 312 AAIACQW--GVWYGF 325

RESULT 30  
US-08-286-767-3  
; Sequence 3, Application US/08286767  
; Patent No. 5733760  
; GENERAL INFORMATION:  
; APPLICANT: Lu, Yichen  
; APPLICANT: Miller, Samuel I.  
; APPLICANT: Killeen, Kevin  
; TITLE OF INVENTION: SALMONELLA VECTORS ENCODING TRUNCATED  
; TITLE OF INVENTION: P9c FUSION PROTEIN, METHOD OF MAKING, AND USES THEREO  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: RONALD I. EISENSTEIN; DIKE, BRONSTEIN,  
; ADDRESSEE: ROBERTS & CUSHMAN  
; STREET: 130 Water Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: US  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/286,767  
; FILING DATE: 05-AUG-1994  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Resnick, David S.  
; REGISTRATION NUMBER: 34235  
; REFERENCE/DOCKET NUMBER: 44420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 523-3400  
; TELEFAX: (617) 523-6400  
; TELEX: 200291 STRE UR  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:

;; LENGTH: 188 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-286-767-3  
Query Match 8.5%; Score 81; DB 1; Length 188;  
Best Local Similarity 23.7%; Pred. No. 0.37;  
Matches 45; Conservative 25; Mismatches 52; Indels 68; Gaps 11;  
QY 32 TDQWAIHYS-YKYTWSRFEKLRG-----KYQDGSYEATFKKEKYKDNRSFA----- 78  
Db 25 TNFASVGVARYAQSKVQDFKRNIRGVNVKYRYEDDSPV-SFISLSLYLYGDRQAGSVEPE 83  
QY 79 -----IGTYSLNDALT--RAGLAYDKAAASKTHLSASIPDTD----- 114  
Db 84 GIHYHDKFEVKYGLMVGFPAYRLSDNFSLYALAGVTYKATEKEH---STQGDSEFSNKI 140  
QY 115 ---RMWSYGATYKFTP---NLSVDVGFHLRGGKKHFEVETQNIKGLLLVEADYTTKATAN 169  
Db 141 SSRKTGFAGAGVQMNPLENIVDVGY-----EGSNIS-----STK--IN 178  
QY 170 LYGLNLNVRF 179  
Db 179 GENVGVGIRF 188

RESULT 31  
US-08-475-989-13  
; Sequence 13, Application US/08475989  
; Patent No. 5679352  
; GENERAL INFORMATION:  
; APPLICANT: CHONG, Pele  
; APPLICANT: KANDIL, Ali  
; APPLICANT: SIA, Charles  
; APPLICANT: KLEIN, Michel  
; TITLE OF INVENTION: Synthetic Haemophilus Influenzae  
; TITLE OF INVENTION: Conjugate Vaccine  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Slim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/475,989  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/256,839  
; FILING DATE: 03-FEB-1993  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA93/00041  
; FILING DATE: 03-FEB-1993  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9202219.3  
; FILING DATE: 03-FEB-1992  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, MICHAEL I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-505 MIS:vg  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155



TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 38 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-475-989-13

Query Match 8.4%; Score 80; DB 1; Length 38;  
Best Local Similarity 75.0%; Pred. No. 0.05;  
Matches 15; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 160 ADYTTKATANLYGLNLYRF 179  
|:|:|:| | | | | | | | | | |  
DB 19 ANYTQAHANLYGLNLYSF 38

RESULT 32  
US-08-475-985-13  
; Sequence 13, Application US/08475985  
; Patent No. 5972349  
; GENERAL INFORMATION:  
; APPLICANT: CHONG, Pele  
; APPLICANT: KANDIL, Ali  
; APPLICANT: SIA, Charles  
; APPLICANT: KLEIN, Michel  
; TITLE OF INVENTION: Synthetic Haemophilus Influenzae  
; TITLE OF INVENTION: Conjugate Vaccine  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION NUMBER: US/08/475,985  
APPLICATION NUMBER: US/08/475,985  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/256,839  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA93/00041  
FILING DATE: 03-FEB-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9202219.3  
FILING DATE: 03-FEB-1992  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, MICHAEL I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-506 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 38 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-475-985-13

Query Match 8.4%; Score 80; DB 2; Length 38;  
Best Local Similarity 75.0%; Pred. No. 0.05;  
Matches 15; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 160 ADYTTKATANLYGLNLYRF 179  
|:|:|:| | | | | | | | | | |  
DB 19 ANYTQAHANLYGLNLYSF 38

RESULT 33  
US-08-472-172-25  
; Sequence 25, Application US/08472172  
; Patent No. 5985288  
; GENERAL INFORMATION:  
; APPLICANT: Munson, Jr., Robert S  
; APPLICANT: Grass, Susan  
; APPLICANT: Chong, Pele Y Y  
; APPLICANT: Fahim, Raafat  
; APPLICANT: SIA, Charles D Y  
; APPLICANT: McVerry, Patrick  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Outer Membrane Protein p1 and Peptides  
; TITLE OF INVENTION: of Haemophilus Influenzae Type B  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R6

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,172  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/849,411  
FILING DATE: 07-JUL-1992  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24973  
REFERENCE/DOCKET NUMBER: 1038-471  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 38 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-472-172-25

Query Match 8.4%; Score 80; DB 2; Length 38;  
Best Local Similarity 75.0%; Pred. No. 0.05;  
Matches 15; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 160 ADYTTKATANLYGLNLYRF 179  
|:|:|:| | | | | | | | | | |  
DB 19 ANYTQAHANLYGLNLYSF 38

RESULT 34  
US-08-256-839-13  
; Sequence 13, Application US/08256839  
; Patent No. 6018019



US-08-256-839-48

```
QY      160 ADYTTKATANLYGLNLNRYF 179
          |:::| | | | | | | |
Db      36 ANYTSQAHANLYGLNLNYSF 55
```

```
QY      160 ADYTTKATANLYGLNLNYRF 179
          |:::| | | | | | | |
Db      36 ANYTSQAHANLYGLNLNYSF 55
```

```

RESULT 39
US-08-475-985-47
: Sequence 47, Application US/08475985
: Patent No. 5972349
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: KANDIL, Ali
: APPLICANT: SIA, Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: Synthetic Haemophilus Influenzae
: TITLE OF INVENTION: Conjugate Vaccine
: NUMBER OF SEQUENCES: 56
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: Suite 701, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/475,985
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/256,839
: FILING DATE: 03-FEB-1993
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/CA93/00041
: FILING DATE: 03-FEB-1993
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9202219.3
: FILING DATE: 03-FEB-1992
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, MICHAEL I.
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-506 MIS:vg
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 47:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 55 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
US-08-475-985-47

```

```

Query Match      8.4%: Score 80; DB 2; Length 55;
Best Local Similarity 75.0%: Pred. No. 0.085;
Matches 15; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 160 ADYTTKATNANLYGLNLYRF 179
:||||:| ||||| |||||
Db 36 ANYTSQAHANLYGLNLYSF 55

```

```

RESULT 40
US-08-256-839-47
: Sequence 47, Application US/08256839
: Patent No. 6018019
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: KANDIL, Ali
: APPLICANT: SIA, Charles
: APPLICANT: KLEIN, Michel

```

```

: TITLE OF INVENTION: Synthetic Haemophilus Influenzae
: TITLE OF INVENTION: Conjugate Vaccine
: NUMBER OF SEQUENCES: 56
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: Suite 701, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/256,839
: FILING DATE:
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, MICHAEL I.
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-373 MIS:jb
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 47:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 55 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
US-08-256-839-47

```

```

Query Match      8.4%: Score 80; DB 3; Length 55;
Best Local Similarity 75.0%: Pred. No. 0.085;
Matches 15; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 160 ADYTTKATNANLYGLNLYRF 179
:||||:| ||||| |||||
Db 36 ANYTSQAHANLYGLNLYSF 55

```

```

Search completed: May 12, 2003, 09:41:59
Job time : 17 secs

```